

Access DB# 84028

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

Jan Delaval
 Reference Librarian
 Biotechnology & Chemical Library
 CM1 1E07 - 703-308-4498
 jan.delaval@uspto.gov

STAFF USE ONLY

Type of Search

Vendors and cost where applicable

Searcher: <u>Jan</u>	NA Sequence (#) _____	STN _____
Searcher Phone #: <u>1448</u>	AA Sequence (#) <input checked="" type="checkbox"/>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>11/12/03</u>	Bibliographic _____	Dr.Link _____
Date Completed: <u>11/12/03</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: _____	Fulltext _____	Sequence Systems <input checked="" type="checkbox"/>
Clerical Prep Time: <u>10</u>	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

181 KMEFFPDQDISPLPQPSRNRKHTALQKLREVNKRLQDLRSCCLSPKHQSPALQSTD 240

```
QY 241 DEVVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
DB 241 DEVVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
QY 301 FGSELSPTATPTTLKLGVAIDICVVLASSSEATTSQELRLRVQGEKQHMLEISLSP 360
DB 301 FGSELSPTATPTTLKLGVAIDICVVLASSSEATTSQELRLRVQGEKQHMLEISLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412
DB 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412

RESULT 2
US-09-175-254-3
; Sequence 3, Application US/09175254
; Patent No. US20020028482A1
; GENERAL INFORMATION:
; APPLICANT: Hong Zhou
; APPLICANT: Jiuqiao Zhao
; APPLICANT: Derong Liu
; TITLE OF INVENTION: NIP45 HUMAN HOMOLOG
; FILE REFERENCE: PHM 70281
; CURRENT APPLICATION NUMBER: US/09/175,254
; CURRENT FILING DATE: 1998-10-20
; EARLIER APPLICATION NUMBER: GB 9722388.7
; EARLIER FILING DATE: 1997-10-24
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-175-254-3
```

```
Query Match 66.7%; Score 1399.5; DB 10; Length 419;
Best Local Similarity 70.1%; Pred. No. 6.8e-93;
Matches 295; Conservative 30; Mismatches 85; Indels 11; Gaps 4;

QY 1 MAEPLRGGRPRSG---GCGARRARGRCPRARQSPARLIPTVLDVLSDSDEVLE 57
DB 1 MAEPLVGRGWGCGAGRGGRGWGGRPRPRAQRSPSGTLDVSVDLVTSDEILLE 60
QY 58 V-----ADPVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEA 111
DB 61 VATARGAADEVEVEPPPPGVFASRDNSGDSGEDRRPAGPPREPVRVRRRLVLDPGEA 120
QY 112 PVVPVYSGKVQSSNLIPDNSSLLKLCPSPEDEADLTNSGSSPSEDDALPSPGWRKLL 171
DB 121 PLVPVYSGKVYSSLRLLIPDLSLLKLYPPGDEBEAEALADSSGLYHEGSPSP-GSPWKTL 179
QY 172 RKCEKEEKMEEPDQDISPLPOPSSRNKSRKHTALQKLREVNKRLQDLRSCLSPKQH 231
DB 180 RTK-DKEEKKKTEFLDNLNGLPSPSPRTKSRTHTRALKLSEVNKRLQDLRSCLSPKPP 238
QY 232 QSPALQSTDEVLVEGVPVLPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 291
DB 239 QGQEQOQOQDEVLVEGVPVLPETPRLPFLKIRCRADLVRLPVRMSEPLQNVVDHMANHLG 298
QY 292 VSPNRILLFGESLSPTATPTTLKLGVAIDICVVLASSSEATTSQELRLRVQGEK 351
DB 299 VSPSRILLFGESLSPTATPTTLKLGVAIDICVVLASSSEATTSQELRLRVQGEK 358

QY 352 QMLSEISLSPSLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVW 411
DB 359 QTLSEVLSRDSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVW 418
QY 412 G 412
DB 419 G 419

RESULT 3
```

```
US-09-925-299-862
; Sequence 862, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 862
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (476)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-862
```

```
Query Match 5.9%; Score 124.5; DB 10; Length 518;
Best Local Similarity 21.4%; Pred. No. 0.13;
Matches 101; Conservative 48; Mismatches 150; Indels 174; Gaps 22;

QY 8 RGRSRGGGARRARGRCPRARQSPARLIPTVLDLV-----SDSD 52
DB 97 RAPR-RGPLGGRRKK---KAPSASDSKADSDGAKPEPVAMARSASSSSSSSSSDSD 151
QY 53 BEVLE-----VADPVEVPVARLPAPAKP-----EQSDSD-----SEGAAGPAGAPRT 96
DB 152 VSVKPPRGRKPAEKPLPKPRGKPKPERPPSSSSSDSDSDSEVDRISEWKRRDEARRREL 211
QY 97 LVRRRR-----RRL-----LDGEPAPVVPVYSGKVQSSNLIPDNSSLLK 136
DB 212 EARRRREOEELRLRLEQEKERRRRADRGEAE-----RSGSGSGDELREDDPEVK 266
QY 137 -----LCPSPED--EADLTNSGSSPSEDDALPSPGWRKLRKCKEKK 181
DB 267 KGRKRGRRGPPSSSDSEPEAELEAKKSAKKPOSSSTEPARKPOKE--KRVPEEKQ 324
QY 182 MEEFPDQDISPLPOPSSRNKSRKHTALQKLREVNKRLQDLRSCLSPKQHSPALQSTDD 241
DB 325 Q-----AKPVKVERTKRSFGSMDRKVEKKE-----PSVEEK--LQKLHS 364
QY 242 EVVLVEGVPVLPQSSRLFTLKI-----RCRADLVRLPVRM--SEPLQNVVDHMANHLGVS 293
DB 365 EI-----KFALKVDSPPVKRCLNLEELGTLTQVTSQILQKNVDVVA----- 405
QY 294 PNRILLFGESLSPTATPTTLK-----LGVADIIDCVVLASSSEATTSQELRLRVQCK 348
DB 406 -----TLKKIRRYKANKDVME-----KAAEVYTLKSKVLGP 437
QY 349 EKHQMLEISLSPSLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLG 401
DB 438 KIEAVQVKNKA-----GMEKEKAEKLAGELAGEEAPQEK 474
```

```
RESULT 4
US-10-052-586-346
; Sequence 346, Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
```

APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C1
CURRENT FILING DATE: 2002-01-15
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059266
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063120
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063121
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063486
PRIOR FILING DATE: 1997-10-21
PRIOR APPLICATION NUMBER: 60/063540
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063541
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063544
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063564
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063734
PRIOR FILING DATE: 1997-10-29
PRIOR APPLICATION NUMBER: 60/063870
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066120
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: 60/066466
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/066772
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/069335
PRIOR FILING DATE: 1997-12-11
PRIOR APPLICATION NUMBER: 60/069425
PRIOR FILING DATE: 1997-12-12
PRIOR APPLICATION NUMBER: 60/069870
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/068017
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/077450
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: 60/077632
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/077649
PRIOR FILING DATE: 1998-03-11
PRIOR APPLICATION NUMBER: 60/078886
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/078939
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079664
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079786
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/080107
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080194
PRIOR FILING DATE: 1998-03-31
PRIOR APPLICATION NUMBER: 60/080327
PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/080333

PRIOR FILING DATE: 1998-04-01
PRIOR APPLICATION NUMBER: 60/081049
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081070
PRIOR FILING DATE: 1998-04-08
PRIOR APPLICATION NUMBER: 60/081195
PRIOR FILING DATE: 1998-04-09
PRIOR APPLICATION NUMBER: 60/081838
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082568
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082569
PRIOR FILING DATE: 1998-04-21
PRIOR APPLICATION NUMBER: 60/082704
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/082797
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/083495
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083496
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083499
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/083559
PRIOR FILING DATE: 1998-04-29
PRIOR APPLICATION NUMBER: 60/084366
PRIOR FILING DATE: 1998-05-05
PRIOR APPLICATION NUMBER: 60/084414
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/084639
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084640
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/084643
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/085573
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085580
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085582
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/085700
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086023
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086486
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/087098
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087208
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05

;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088722
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088740
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088811
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088825
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088863
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089090
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 5.9%; Score 124.5; DB 12; Length 671;
Best Local Similarity 23.1%; Pred. No. 0.19;
Matches 93; Conservative 41; Mismatches 143; Indels 125; Gaps 20;
QY 8 RGRPRSGRGARRARGRCPRARQSPARLIPDTVLVDLV-----SDSD 52
Db 213 RAR-RGPLGRKKK----KAPASDSKADSDGAKPEPVAMARSASSSSSSSSSDSD 267
QY 53 EEYLE-----VADPVEPVABLPAPAKP-----EQDSQD-----SEGAAGPAGAPRT 96
Db 268 VSVKPPRGRKPAEKPLPKRGRKPKPERPPSSSSSDSDSDVDVIRSEWKRRDEARRREL 327
QY 97 LVRRRR-----RL-----LDGPEAPVPVYSGKVOSLLNLPDSSLK 136
Db 328 EARRRREQEELRLREKEKEKERRERADRGAE-----RSGGSSGDELREDDEPVK 392
QY 137 -----LCSPSEPD--EADITNSGSSPSDDALPGSPWRKKLR----- 172
Db 383 KRGRKGRGCRPPSSDSEPAELERAKKSAKXPQSSSTEPARKPGQKEKVRPEEKQQA 442
QY 173 --KKCKEKKMBEFP--DQDISPLPOPSSRNKRKTEALQKLR--EVNKKLQ-----DLRS 224
Db 443 KPVKVERTKRSEGFSDMRKVEKKEKPSVEEK-----LQKLHSEIKFALKVDSPDKR 495
QY 225 CLS-----PKHQSPALQSTDDVNLVE-----GPVLPOSSRLFTLKIRCRADLV 270
Db 496 CLNALBELGLQVTSQLQKNTDVATLKIRRYKANKOVNKAEEVYT-----R 545

QY 271 LPVRMSEPLQNVVDHMANHLGVSPNRILLLLFGESELSPTATP 312
Db 546 LKSRVLGPKIEAVQKV-NKAGMEKEKABEKLAGEELAGEEAP 586
RESULT 5
US-09-823-356-9
; Sequence 9, Application US/09823356
; Patent No. US20010025098A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Baughn, Mariah R.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN MEMBRANE SPANNING PROTEINS
; FILE REFERENCE: PP-0489-1 CON
; CURRENT APPLICATION NUMBER: US/09/823,356
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/039,307
; PRIOR FILING DATE: 1998 March 13
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PERL Program
; SEQ ID NO 9
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010025098A1 1794154
US-09-823-356-9

Query Match 5.9%; Score 123; DB 10; Length 950;
Best Local Similarity 19.9%; Pred. No. 0.38;
Matches 108; Conservative 74; Mismatches 203; Indels 158; Gaps 23;
QY 3 EPLRGGRPRSGRGARRARGRCPRARQSP-----ARLIPDTVLVDLVSDSDDEVLEV 58
Db 314 DPRRAQVPPPTRESR-VKVLRLAATAPALPOPPSTPRATTLPPAPATTVTRTSRAVTV 372
QY 59 ADPV-----EVPVARLPA-----PAKPEQSDS----- 80
Db 373 ARPMTTTPPTTQRTWTPSPSHRPPTTTEVITARRPSVSENLYPPSRKDDQHRERPTTTR 432
QY 81 SDSEGAAGPAGAPRTLVRRRRRRLDPG-----EAPVVPVYSGKVQSSL 125
Db 433 PSKATSLSEFTNAPTITISEPSTRAAGPGRFDRNDRMRHGHGRDPNVVP---GPPKPAK 489
QY 126 NLIPDNSSLKLCPSPEDEADLTNSGSSPSDDALPGSPWRK-----KLKKCKEK 177
Db 490 EKPAPKKAQDKILSEYBEKYDLSRPTASQLEDELQVGNVPLKKAKESKKHEKLEKEPE 549
QY 178 EKKME-BFPDODISPLPOPSSRNKRKHTEALQKLREVNKLQD----- 221
Db 550 KKKKMKENADKLLKSEKQMKKSEKSKQKEKSKKGGKTEQDQGYQKPTNKHFTQSPK 609
QY 222 -----LRSLSPKQHSQSPALQSTDD-----EVLVEGVPVLPQ 253
Db 610 KSVADLLGSFEGKRRLLLTITAPKAENMVMVQQRDEYLESFCKWATRKISVITFGPV--- 666
QY 254 SSRLLFTLKIRCRADLVRL-----PVRM--SEPL--QNVVDHMANHLGVSPNRILLIFGESE 305
Db 667 --NNSTMKI-----DHFQLDNEKPMRVDDDELVDQRLISLSEKRYGMYTNDYFVMTD 720
QY 306 LSPT---ATPSTLKLGVADIICVVVLASSSEATETSOELRLRVQGE-KHOMLETSISP- 360
Db 721 LVKQYVEVPIIMK-SVFDLID-----TFQSRIDKMEKQKKEGIVCKEDKKQSLFNLSRF 775

QY 361 -----DPLKVMGSHYEAMGLSGHLSF---FPGTKLSKELPADLGLSGSLI 408
DB 776 RWRRLVLVISAENDWMAYSQOLSLSGQACNFGRLHITILKLG-----VGEVGVGL 829
QY 409 EVW 411
DB 830 ELF 832

RESULT 6

US-09-738-877-3
Sequence 3, Application US/09738877
Patent No. US20020015970A1
GENERAL INFORMATION:
APPLICANT: Murray, Richard
APPLICANT: Watson, Susan
APPLICANT: Weises, Stephen J.
APPLICANT: Glynn, Richard
TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
FILE REFERENCE: A-69806/DJB/JJD
CURRENT APPLICATION NUMBER: US/09/738,877
PRIOR FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: US 09/637,977
PRIOR FILING DATE: 2000-08-11
PRIOR APPLICATION NUMBER: US 60/148,425
PRIOR FILING DATE: 1999-08-11
PRIOR APPLICATION NUMBER: PCT/US 00/22061
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent version 3.1
SEQ ID NO: 3
LENGTH: 1781
TYPE: PRT
ORGANISM: Homo sapiens
US-09-738-877-3

Query Match 5.8%; Score 122.5; DB 10; Length 1781;
Best Local Similarity 22.0%; Pred. No. 0.92;
Matches 93; Conservative 62; Mismatches 152; Indels 115; Gaps 22;

QY 40 PPTV-LVDIVSDSEVLEVADPVEVPVARLPAPAK---PEQSDSDSEGAAGPAGAR 95
DB 192 PPTVOLLTVKKEGEGAGAGADHODPSLGAAGEAASKESPEKOSTEKPE-----240
QY 96 TLVRRRRRLDPCGAPVVPVYSGKVQSSLNLPDSSLLKLCPSPEP--EADLTNSGS 153
DB 241 TLKRQSHAEISP-----PAESGA-----VEECKEKEKEKEPEKSAE 281
QY 154 SPSEDDALPSGSP-----WRKK--LRKCKEKE---EKKMEPEPDODISPLPQ 196
DB 282 SPTSIVTSTGTFKFFTOGAGNRKKTSPFKPKDEVEASEKKKEQPEK-----VDT 336
QY 197 SSRNRSRKATEALQKLRVYNKRLQDLRSLSPK---QHOSPALQSTDEVLVEGP---249
DB 337 EEDGAEEVASEKLTASEQHQP--EPASAHBEPRLSAEYKVELPS--EEGVSSQSGSEK 394
QY 250 VLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDMANHLGVSPNRILLFGESLSPT 309
DB 395 PAPLATEVDEKIEVHOEEVVAHVSTVEERTEQ-----KTEVEERT 437
QY 310 A--TSTTLKLGVAADIICVVLASSSEATETSOELRLR---VQKKEGQMLLEISLSPSL 364
DB 438 AGSVAEAEIVGND-----AEPOEAEPAKELVKIKETCVSGEDTQAD--LSDEKV 487
QY 365 -----KVLMSHYEBAMGLSGHLSFPDGT---KLSGKELPADLG---LESGLI 408
DB 488 LSKPEGVVSEVEMLSQGRMKVQSGPLKLTFTSTGLKTLGKQKQKRGGGBESGERT 547
QY 409 EV 410
DB 548 QV 549

RESULT 7

US-09-765-272-218
Sequence 218, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 218:
SEQUENCE CHARACTERISTICS:
LENGTH: 565 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-09-765-272-218

Query Match 5.7%; Score 120.5; DB 10; Length 565;
Best Local Similarity 20.9%; Pred. No. 0.29;
Matches 81; Conservative 66; Mismatches 156; Indels 85; Gaps 19;

QY 29 PRAROSP-----ARLIPPTVLVDIVSDSEVLEVADPVEV--PVARLP---APAKPE-Q 77
DB 34 PTEESPKEEPKSEVKKPDDTLPKVEBGKEDSABPAVVEVGGEVSKPEEKVAVKPESQ 93
QY 78 DSDSDSEGAAGPAGAPPTLVRRRRRLDPCGAPVVPVYSGKVQSS--LNLIPDSSSL 134
DB 94 PSDKRAESKYEQAEP--VAREDEKAPVEPEKQPEAEBAVEETPKQESTDTYKAE 152
QY 135 LKLCP-----SEPEDEADLTNSGS--SPSEDDALPSGSPWRKK 170
DB 153 ETVPEKEETVANSIQPKVETPAVEKQTEPEEPKVEQAGPEVAREDEQAFT--APEPE 211
QY 171 LRKKEKEKEKEEPPDODISPLPOSSRNRSRKATEALQKLRVYNKRLQDLRSLSPKQ 230
DB 212 KOPEVPEEKAVEE-----TPKPEKIKIGTKEPVDR--SEINNOI--DRASSVSPTD 261
QY 231 HOSPALQSTDEVLVEGPVLPPOSSRLFTLKIRCRADLVRLPVRMSE--PLQNVVD---284
DB 262 YSTASYNL-----GPVLETAKGYA-----SEPVKQEVNSETNKLKLTALDALNV 307
QY 285 ---HMANHLGVSPNRILLFGESLSPTATPSTLKLGVADIIICVVLASSSEATETSOEL 341
DB 308 DKTELNTIADAKTKVKEHYSDRSWONLQTEVTYKAEKVAANTD---AKOSEVNEAVEKLT 363

```
QY 342 RLAVOGKQKHMLEISLSPDPLKVLMS 369
Db 364 TATIE-----KLVELS---EXPILTITS 383

RESULT 8
US-09-931-087A-1
; Sequence 1, Application US/09931087A
; Patent No. US20020147322A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: NSP Molecules
; FILE REFERENCE: P1223RIE
; CURRENT APPLICATION NUMBER: US/09/931,087A
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/367,206
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/082,767
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 35
; SEQ ID NO 1
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-09-931-087A-1

Query Match 5.6%; Score 118; DB 10; Length 576;
Best Local Similarity 21.8%; Pred. No. 0.45;
Matches 90; Conservative 49; Mismatches 148; Indels 126; Gaps 18;

QY 3 EPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLSDSDDEVLVADPV 62
Db 137 EPLRAR-----KWSNQPADLAHMGSRSDPDAGMEASTMPISALPRTS-----SDPV 183

QY 63 EYPVARLPAPAKPEQDS--DSDSEGAAGPAGAPRTLVRRRRRRLDPGEAPVVPVYSGK 120
Db 184 ---LLKAPAPLGTVDASLRASDQLOKAKTPKPRP-----PSFELPDASERP--PTY--- 231

QY 121 VOSSLNLIPDNSSLLKCPSEPEDEADLTNSGSSPSDEDDALPSGSPWRKKLRKCKEKEK 180
Db 232 -----CELVRPVPSVQ-----GTSPSQSCPEPE-APWWE--AEDEEEN 268

QY 181 KMEFPDQDISPLPQSPSRNKRKHTALQKLRVKNKLDLRSCLSPKQHQ--SPALQS 238
Db 269 RCFTRPQAEISFCP-----HDAPSCLLGPQNRPLEPOVLH 303

QY 239 TDDEVVLVEGPVL---POSSRLFTLKIRCA-----DLVRLPVRM 275
Db 304 T-----LRGLFLEHPGSTALHLLVDQATGLLVTRDQGRNMGVSSGLELLTLPHG 357

QY 276 SEPLQNVVDHMANHL-----GVSPNRILLFGESELSPTATPSTL--KLGVADI 323
Db 358 HRLLELHERHOTLALAGALVLCGSPLEERAAALRGVLALALRPGNAGDPLGLAAYM 417

QY 324 DCVVLASSEATTSQELRLRVQKQKHMLEISLSPDPLKVLMSHYEEAMG 376
Db 418 GALLMPQVSRLEHTWRQLR-----RSHTEAALAFEQELKPLMRALDEGAG 462

RESULT 9
US-09-931-087A-21
; Sequence 21, Application US/09931087A
; Patent No. US20020147322A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: NSP Molecules
; FILE REFERENCE: P1223RIE
; CURRENT APPLICATION NUMBER: US/09/931,087A
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/367,206
```


Best Local Similarity 21.8%; Pred. No. 0.45;
Matches 90; Conservative 49; Mismatches 148; Indels 126; Gaps 18

QY	3	EPLEGRGRSRNGSGGARARARARARGRGCRPARROSAPALITDVTLVLDVVSDEVEVLVADPV	62
Db	137	EPLERR-----KWSNQPADLHMGSRSDPAGMEASTYPIALPRTS-----SDPV	183
QY	63	EVVAPARLPAPAKPEODS---DSDSEGAAGPAGAPATLVRRRRRLIDPGEAPVVPVYSGK	120
Db	184	---LLKAPRLGYADSLRASDGLQAKAPTKPRT---PSFELPDASENP---PTY---	233
QY	121	VOSSLNLIPDNSSLKLCPSEPEDEADLTNGSSPESD DALPSSGPMWRKLKKCEKEK	180
Db	232	-----CELVRVPVSQ-----GTSPSGCSPEPE-APWME-AEEDBEEN	266
QY	181	KMEERPDODISPLRQPSRRNRKRTTELQKLRVKNKRLDRLRGLSLCKQHQ--SPALQS	233
Db	269	RCFTRPQAEISFCP-----HAPSCLLGPORRPLEPVLH	303
QY	239	TDDEVVLVEGVUL-----POSSRLFTLTIKRCRA-----DLVRILPRM	275
Db	304	T-----LRGLFLEHNPRSTALHLLLVNCCQTGLLGYTRDRGMMGVSSGIELTLTPRGH	355
QY	276	SEPLQNVVDHMANH-----GVSNNRILLFGSESLSPATPSTL--KLGVADI	322
Db	358	HLRLELERRHQTALAGALAVLGGSGPLREBRAAALRGVELALALRPGADDLRGLAAM	411
QY	324	DCVVLASSELETTSQELRLKVGCKEKQMEISLSPSPKLVMSHYEAMG	376
Db	418	GALLMPQVSRLEHTWRQLR-----RSHTEALAFPEQELKPLMRALDEAG	462

```

RESULT 11
US-09-931-087A-23
; Sequence 23, Application US/09931087A
; Patent No. US2002014732A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; TITLE OF INVENTION: Nsp Molecules
; FILE REFERENCE: P1223RIE
; CURRENT APPLICATION NUMBER: US/09/931,087A
; CURRENT FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/367,206
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 60/082,767
; PRIOR FILING DATE: 1998-04-23
; PRIOR APPLICATION NUMBER: US 60/113,296
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 35
;
; SEQ ID NO 23
;
; LENGTH: 576
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
;
; OTHER INFORMATION: Mutation of SEQ ID NO.11:US-09-931-087A-23

```

Query Match	5.5%;	Score 115;	DB 10;	Length 576;
Best Local Similarity	21.1%;	Pred. No. 0.74;		
Matches	88;	Conservative	52;	Mismatches 143;
				Indels 134;
				Gaps 18;

QY	3	EPLTGRGPRSGGAGRRARAGRCPRAROSPRLI	PDITLVLTVDSDSEVLELVADPV	62
Db	137	EPLTAR-----KMSNSQPADLHHMGRSRDPAGME	STIPALRPTS-----SDPV	183
QY	63	EVPAFLRLPAKPKPEQDS--DSDSCGAAGPGAGP	RTLYRRRRRLDQGAENVTVYSGK	120
Db	184	---LKKPAFLPAGTVADISRASDGLQAAAPKKPPT	-----PSE--	220
QY	121	VOSSLNLTIPDNS---SLTKLCPSEPEDEADVT	NSGSSPSEEDALPQSGSPWRKTLRKCE	176
Db	221	---LPPDASRPPTFCCLVPPVPSVQ-----	GPSGSCSEPE-APWME--AAEDE	264

Gy	177	KEKKMEERFPRDII SPLRQSSRNKSKKHTEALQKLEAVNKRIQDLRSGLSPRONO--SP	234
Db	265	EENKCTFRPAETISFCP-----HDAPSCILGPNRRLEP	299
Gy	235	ALOSTDEVLVVEEPPVL--POSSRYLFTLKRCRA	-DLVRL 271
Db	300	QLVLT-----LRRLFBNHNGSTALLLYLVDCQTSLGVTRDQRGNMVS SSGLELLT	353
Gy	272	PVRNSEPLQNVDHMANHL-----GVSPNRILLFGESLSPTATPSTL--KLGV	319
Db	354	PGHNLRLELEHERQTLALGALVALTCGSPGLEBAARALLGLVELMALALPCGAAGDLPGL	413
Gy	320	ADIIDCVLAASSATETSQELRIYVGCKKKNQMLETSISPDSRPKYLMSHYEAMC	376
Db	414	AADVALLMPQVSLERTKQIR-----RGHTAEALAFFEDCLKYMLDALDEGAG	462

```

: RESULT 12
: US-09-864-761-36307
: Sequence 36307; Application US/09864761
: Patent No. US2002048763A1
: GENERAL INFORMATION:
:   APPLICANT: Penn, Sharon G.
:   APPLICANT: Rank, David R.
:   APPLICANT: Hanzel, David K.
:   APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecmica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annumax Sequence Listing Engine vers. 1.1
: SEQ ID NO 36307
: LENGTH: 92
: TYPE: PRT
: ORGANISM: Homo sapiens

```


[illegible]

Qy	6	RGRGRSRSGGGARRRARRGRCRCRAQSPARKLIPDTVLVDVSDSDBEVLVADPVEVP	65
Db	455	RQRGGKQGGSSPQRASG-----FNRHQAPSLSPCPA--LSSDSKEGEDECTEELP	506
Qy	66	-----VARLAPAKPEDDDSDSDSGAE-----GAGCAPRTIVRRRRRL	106
Db	507	ALPYLAKSTKKALSVSPSPALPRSLSIHWSMRAQESVGLDPAPAAANPCP---RRGRWV	563
Qy	107	DGEAPVIVPVYSGVQSSLNLIPINSSLTKLCPS--RP-EDENLUTNGSSPSSEDDLPS	166
Db	564	QPG-----VELSVSMIDLROLLETAPSLQDPSPQSSIALIPSGPRKIQEALET	612
Qy	164	G-----SPWKKLRRKC-----EKEE-----KKMEEPDDDISPLPOPS--SRNK	201
Db	613	SLTSQNEKPPRPAQSPQSPYHIIIRLLSQEBGVAFQDLERAPLEDGIVYPEPSONPMOT	677
Qy	202	SRKRTKALQK-----LREVNKRLLQD-----LRCSLCPKQHOSPALQSTDDEVL	245
Db	673	SEFQVQAPARCTLGRVYFGSRSSSKHSPDASCVDVSSCLSSRH--PRDESSSTEPLS	733
Qy	246	VEGEVLPOSSRLFTLKIRCRADIVLRPEVMSSEPLQNVYDHANHLGVSPNRILLPGESE	305
Db	731	VDG-----ISSDLEPRAEQDEEBEEREGNGP-----YGLQE	766
Qy	306	LSPTATP-----STLKLGVADIIDCVVLASSSEATETSGELRLVQGEKKHOMLE	355
Db	763	GSP-QTDPDQEQFLQKHETLTLASGAAPAP--VQYPERSESSISSRFLQYQTRLRAP--	818
Qy	356	ISLSPDSPFLKVLMSHYEAMGLSGHK	381

Db 819 ---SPSSSLALMSRPAQVPOASGEQ 841

RESULT 2

A37221

neurofilament triplet H protein - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999

C:Accession: A37221; A25649; A30796; A32757; B25649

R:Chin, S.S.M.; Liem, R.K.H.

J:Neurosci. 10, 3714-3726, 1990

A:Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vimentin

A:Reference number: A37221; MUID:91038277; PMID:2230956

A:Accession: A37221

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1072 <CHI>

A:Cross-references: GB:AF031879; NID:g2642597; PIDN:AAB87068.1; PID:g2642598

R:Robinson, P.A.; Wion, D.; Anderton, B.H.

FEBS Lett. 209, 203-205, 1986

A:Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).

A:Reference number: A25649; MUID:87080760; PMID:2878828

A:Accession: A25649

A:Molecule type: mRNA

A:Residues: 230-318:472-542 <ROB>

A:Cross-references: GB:M37327

R:Dautigny, A.; Phan-Dinh, D.; Roussel, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.

Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988

A:Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ detection

A:Reference number: A30796; MUID:88309090; PMID:2457365

A:Accession: A30796

A:Molecule type: mRNA

A:Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-731

A:Cross-references: GB:M21964; NID:g205685; PIDN:AAA41695.1; PID:g205686

R:Lieberburg, I.; Spinner, N.; Snyder, S.; Anderson, J.; Goldgaber, D.; Smulowitz, M.; C

Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989

A:Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide

A:Reference number: A32757; MUID:89184647; PMID:2928342

A:Accession: A32757

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 559-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <LIE>

A:Cross-references: GB:J04517; NID:g205679; PIDN:AAA41692.1; PID:g205680

C:Superfamily: neurofilament triplet H protein

C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 6.9%; Score 145.5; DB 1; Length 1072;

Best Local Similarity 22.1%; Pred. No. 0.33;

Matches 95; Conservative 63; Mismatches 166; Indels 105; Gaps 20;

QY 4 PLRGPRSGRGARRGARGRCPRAROSPARL-IPDVLVLDVSDSEEV----- 55

Db 613 PVAKSPAEAKSPASVKSPPG-EAKSPAEAKSPAEVKSPTATKSPAEAKSPVTVK 671

QY 56 --LEVADPEVVPVAPLAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEAPV 113

Db 672 SPAEAKSPVEV-----KSPASVKSPEAKSPAGAKSPAEAKSPVPAKSPAEAK- 725

QY 114 VPVYSGVQSGLNLI PNSSLLKLCPSPEDEADLTN-----SGSSPSDDA---LP 162

Db 726 -PAAEAKSPAEAKSPAEAKS-----PAAEAKSPAEAKSPVEKSPAEKSPVKEGAKSLAE 779

QY 163 SGSPWRKLRKKCKEKKMEFPDQDISPLPQPS-SRNKSR-----KHTALQKLRVNVK 217

Db 780 AKSPAEAKSPVEIKPEAEVKSPEAKSPWKEAKSPERAKTLVDKSPAEKTPAKEEAK 839

QY 218 RLQDLRSLSPKQHQSPALQSTQDDEVVLVVEGPNLPQSSRLFTLKIRCRADLVRLP---VR 274

Db 840 RPADIR---SPEQVKSPEAKEAKS-----PEKEETRTKVPAPKEEVKSPVEEVK 866

QY 275 MSEPLQNVVDHMANHLGVSPNRIILLFGSELSPTATPSTLKLGVDIICVVLASSSEA 334

Db 887 AKEP-----PKV-----EEKTP-ATPKTE-----VKSKKQDA 915

QY 335 TETSOELRLRVQKQKQHMLEISLSPDPLKVLMSHYEAMGLSGHKLKLSFFFDGTLKSK 394

Db 916 PKEAQ-----KPKAEEKEPLTE--KPKDSPGEAKKEEAKKAAPE-----E 956

QY 395 ELPADLGL 403

Db 957 ETPAKLGVK 965

RESULT 3

T26517

hypothetical protein Y18D10A.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26517

R:Harris, B.

submitted to the EMBL Data Library, December 1998

A:Reference number: 220226

A:Accession: T26517

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1634 <WIL>

A:Cross-references: EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1

C:Experimental source: clone Y18D10A

C:Genetics:

A:Gene: CESP:Y18D10A.1

A:Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2;

Query Match 6.7%; Score 140; DB 2; Length 1634;

Best Local Similarity 22.0%; Pred. No. 1.2;

Matches 88; Conservative 54; Mismatches 152; Indels 106; Gaps 18;

QY 3 EPLRGPRSGRGARRGARGRCPRAROSPARLIPDVLVLDVSDSEEVLEVADP- 61

Db 413 EMKGESVKEKAARWRAEASAGRSQAQGPAPAPA-----ASELODPPQDFGLSMSDPG 465

QY 62 -----VEVPV-----ARLPA-PAKPEQSDSDSEGAAGPAGAPRTLVRRRR 102

Db 466 SDSETEEHQKQKHIPAVMYTRRSARLSALPVTPKKASSSK--MPPPSPSTPCRRGR 523

QY 103 R-----RLLDGCEAPVPVYSGKVQS-SLNLI PNSSLLKLCPSPEDEA-DLTNSGSS 154

Db 524 RPRTLSTMSMEPAAAVTPAPRGPRSRSAKVSENTEPLSEAPAPVKRGRGRSRST 583

QY 155 PS-EDDALPSGSPWRKLRKKCKEKKMEFPDQDISPLPQPSRNKSRKHT--EALOK 211

Db 584 MSITEDSEPTSTSTAAKRSKRAEEDDEEODLJLTKNSP-EKPKPKSTTEETVGDVLKK 642

QY 212 -LREVNK-----RLQDLRSLSPKQHQSPALQSTD-----DE 242

Db 643 RLRLDTAKTTATVIHTPGPLTRTKMRAPTAVTSSKKEKPNAGSADSSINEEHEDE 702

QY 243 VVLVEGVP--LPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPNRIILL 300

Db 703 TWILEBQTLDLPOQTSQBPRIKSGSEL-----LDEQFDASEHSGTVPS----- 747

QY 301 FGESELSPTATPSTLKLGVDIICVVLASSSEATETSQ 340

Db 748 --APELTKNPAPV-----PEASEASAE 768

RESULT 4

S02003

neurofilament triplet H protein - rat (fragment)

C:Species: Rattus norvegicus (Norway rat)

C>Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 13-Aug-1999

C:Accession: S02003

R:Breen, K.C.; Robinson, P.A.; Wion, D.; Anderton, B.H.

FEBS Lett. 241, 213-218, 1988

A:Title: Partial sequence of the rat heavy neurofilament polypeptide (NF-H). Identifica

A:Reference number: S02003; MUID:89065087; PMID:3143606
 A:Accession: S02003
 A:Molecule type: mRNA
 A:Residues: 1-854 <BRE>
 A:Cross-references: EMBL:X13804; NID:957828; PIDN:CA32038.1; PID:957829
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil

Query Match 6.6%; Score 137.5; DB 2; Length 854;
 Best Local Similarity 21.9%; Pred. No. 0.77;
 Matches 94; Conservative 62; Mismatches 168; Indels 105; Gaps 20;

QY 4 PLRGPRGRCGRARRARGRCPRARQSPARL-IDTVLVLDVSDSEV----- 55
 DB 372 PVEAKSPAEAKSPASVSKPG--EAKSPAEAKSPAEVSPATVSPAEAKSPAEVSPATV 430
 DB 431 SPAEAKSEVEY-----KSPASVSPSEAKSPAGAKSPAEAKSPVAKSPAEAKSPAGAK- 484
 QY 114 VPVYSGVQSSLNLPDSSSLKLCPSPEDEADLTN-----SGSSPEDEDA---LP 162
 DB 485 -PPEAKSPAEAKSPAEAKS-----PAKSPAEAKSPAEVSPAEKSPAEKSKSLAE 538
 QY 163 SGSPWRKKLRRKCEKEKMEFPDQDISPL-PPSSRNKSR---KHTALQKLRVNVK 217
 DB 539 AKSPEKAKSPVKEIKPPAEVSPAEKSPAEKSPAEKSKTLDVKSPEAKPPAEK 598
 QY 218 RLQDLRSLSKQKQSPALQSTDEEVVLVEGVLPOSSRLTLTKRCADLVRLP---VR 274
 DB 599 RPADIR--SEQVKSPEAKSEAKS-----PEKEETREKVPKKEEVKSEVSEVK 645
 QY 275 MSEPLQNVDMANHLGVSPRILLFGESLSPATPSTIKLGADIIDCVVLASSSEA 334
 DB 646 AKEP-----PKV-----EEKTP-ATPKT-----VKESKDEA 674
 QY 335 TETSOELRLRVQGEKQKQMLISLSPDSPLVLMSHYEAMGLSGHKLSPFDGTXLSGK 394
 DB 675 PKAQ-----KKAEKEPLTE--KPKDSPGEAKKEAKEKAAABE-----E 715
 QY 395 ELPADLGLG 403
 DB 716 ETPAKLGVK 724

RESULT 5

WMBETE

65K early nonstructural protein - human cytomegalovirus (strain Towne)
 C:Species: human cytomegalovirus, human herpesvirus 5
 A:Note: host Homo sapiens (man)
 C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C:Accession: A41808
 R:He, Y.S., Xu, L., Huang, E.S.
 J. Virol. 66, 1098-1108, 1992
 A>Title: Characterization of human cytomegalovirus UL84 early gene and identification of
 A:Reference number: A41808; MUID:92114132; PMID:1309892
 A:Accession: A41808
 A:Molecule type: mRNA
 A:Residues: 1-587 <HEV>
 A:Cross-references: GB:M81432; NID:9330541; PIDN:AAA45947.1; PID:9330542
 C:Superfamily: human cytomegalovirus 65K early nonstructural protein
 C:Keywords: leucine zipper; nonstructural protein
 F:114-135/Region: leucine zipper motif
 F:325-373/Region: leucine zipper motif

Query Match

6.4%; Score 134.5; DB 1; Length 587;
 Best Local Similarity 24.4%; Pred. No. 0.74;
 Matches 105; Conservative 58; Mismatches 159; Indels 109; Gaps 22;

QY 6 RGRPRGR-----GGRARRARGRCPRARQSPARLIDTVLVLDL--VSDSDEV 55
 DB 11 RARRPRARRGGGGVGSNSRRS--GKCRRRARRALSAPLTLATTTTTTMMGVASTDDSD 69

QY 56 LEVADPEVEVVARLPAPAKPEQSDSDSEGAEGPAGPRTLVRRRRLDPGEAPVP 115
 DB 70 LLLKTPDEL-----DKHSGSPQTL-----TLTDKHDIRQPR 101
 QY 116 VYSGK---VQSSLNLPDN-----SSLKLCPSPEDEADLTNCGSSPEDEADLPS 163
 DB 102 VHRGTYHLIQHLDLRPELRDPQIILLSTPLQGEANGESO--TADATSOEETLA-- 156
 QY 164 GSPWRKKLRRKCEKEKMEFPDQDISPLPPSSRNKSRKHTMALQKLRVNVKLDLDR 223
 DB 157 -----SHELKKEKEKEKEE--DED-----DRNDRER-----GLCVANSNDSDVR 198
 QY 224 SCLG--PKQHSFALQSTDEEVVLVEGVLPOSSRLFTLTKRCADLVRLPYRMSEPLON 281
 DB 199 PAFSLFPRPCCHILRSVIDQ--QLTRMAIVRLSLNLFALRI--ITPPLKRVPLRRKAHHT 256
 QY 282 VV--DHMANHLGVSPRILLFGESLSPATP--PSTLKL-GVADIIDCVLASSSATETS 338
 DB 257 ALHDCMALHL-----PELTPESTLIDINNVTENASVADAESTDADLT 299
 QY 339 QELRLRVQGEKQKQMLISLSPDSPLVK-----LMSHYEAMGLSGHKLSPFDGTXLSG 393
 DB 300 PTLTVRVNHAQWHRVEGGISGPRGLTISRISARLSETTAKTLGFS-----VFGRLDLP 353
 QY 394 KELPADLGLG 404
 DB 354 NESPPDLTSS 364

RESULT 6

hypochemical

protein C25A1.10 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T19431
 R:Mortimore, B.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19124
 A:Accession: T19431
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-971 <WIL>
 A:Cross-references: EMBL:Z81038; PIDN:CAB02755.1; GSPDB:GN00019; CESP:C25A1.10
 A:Experimental source: clone C25A1
 C:Genetics:
 A:Gene: CESP:C25A1.10
 A:Map position: 1
 A:Insertions: 38/3; 92/3; 201/3; 919/3
 C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein

Query Match 6.3%; Score 133; DB 2; Length 971;
 Best Local Similarity 23.3%; Pred. No. 1.7;
 Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;

QY 29 PRARQSPARLIDTVLVLDVLS--DSDEVELEVADPEVVARLPAPAKPEQSDSDSEG 85
 DB 231 PAKRTPAKRTPKPVVKKAESSSDSDDEKKPVAKP--APAKATPKPAKKAADSSSDSD 288
 QY 86 ---AEGPAGAPRTLVRRRRLLD-----PGEAPVPVYSGVQSSLNLI 128
 DB 289 DEAPAKTPAKAARPVVAKKAESSSDSDDEKKPAKTPAKATPKPVAKKAESSSDSD 348
 QY 129 PDNSSLKLCPSF--PDEADLTNCGSSPSRDDALPSSSP--RKLTKRKCEKEKMEEF 185
 DB 349 DEKKPVAKPAKATPKPVAKKAESSSDSDDEKKPAKTPAKATPKPVAKKAESSSDS 408
 QY 186 PDQDISPLPPSSRNKS-----RKHTALQKLRVNVKRLQDLRSC 225
 DB 409 SDDEKKVPKATPKATPKPAKKAADSSSDSDDEAPAKTPAKAARPAKKAAL---SS 465
 QY 226 LSPKQHSFALQSTDEV 243
 DB 466 SDSDDEKKPAKSTPAKI 483

RESULT 7

T29187
hypothetical protein C55C3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R:Accession: T29187
R:Woessne, J.; Stellyes, L.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid C55C3.
A:Reference number: Z20585
A:Accession: T29187
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-792 <WOE>
A:Cross-references: EMBL:U53335; PIDN:AAA96170.1; GSPDB:GN00022; CESP:C55C3.3
A:Experimental source: strain Bristol N2; clone C55C3
C:Genetics:
A:Map position: 4
A:Introns: 17/1; 46/3; 139/2; 173/1; 204/3; 250/3; 266/3; 286/2; 327/1; 362/3; 393/3; 43

Query Match 6.3%; Score 132.5; DB 2; Length 792;
Best Local Similarity 20.9%; Pred. No. 1.4;
Matches 90; Conservative 50; Mismatches 135; Indels 155; Gaps 19;
QY 11 RSGRGARRARGRCRCRPAROSPARLIPDTVLVDLSDSEVLEVADPVEVPVARLP 70
DB 309 RRGGNQIKKKPAK---PLKRITP-----NNSDKLKMKSP----- 343
QY 71 APAPKQDSDS-----SEGAAEGPAGAPRTLVRRRRR-----LLDPG 109
DB 344 -PKPLQATSDILQPLPSNNAMTVPEKESANRSPQSSKINLVNRGRTVNTDLKDCS 402
QY 110 EAPVVPYSGKVOSSLNLPDNLKCPSEPEDEADLTNSGSSPSEDALPSGSPWK 169
DB 403 EARAIP-----KC-DKLTDFQTSTNSQSPA--DAISAVAP-TK 437
QY 170 KLRKKEKEKKMEEPD-----QDI-----SPLQPSRKN----- 201
DB 438 LTGSQSQSERQIEKMPNSFRELQDSKFKPAEAPIKPEERKSGTEQSTTEPTKLGSQ 497
QY 202 SRKHTALQKLRVKNLQD--LRSCLSPKHQSPALQSTDD----- 241
DB 498 SPSEQEIQQMPNSFRELQDSKFKPAPSPKSEKPAEQSLSESPSPVPGNNRRNSETSQVD 557
QY 242 -----EVLVGGPVLQPSRSLFTKIRCADLVLPLVRMSEPLQNVVDH-----MAN 288
DB 558 TISPVPTKLVTQSPSERQIEQMPNSFRELQDSKFKPAQAPNPKPERVERGSAEQSMSE 617
QY 289 HLGVPNRIILLFGESLSPTATPSTLK-----LGVADIIDCVILA-----SS 331
DB 618 PLIS-----RVAG-SPIAPKPRSPPLQAPLLETATPTPIDAPTAAIETASAEFSS 672
QY 332 SEATETSQEL 341
DB 673 SHSMDPSNSL 682

RESULT 8

WMBEDE
65K early nonstructural protein - human cytomegalovirus (strain AD169)
N:Alternate names: UL84 protein
C:Species: human cytomegalovirus, human herpesvirus 5
A:Note: host Homo sapiens (man)
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
C:Accession: S09848
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Horsnell, T.;
M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
A:Reference number: S09749; MUID:90269039; PMID:2161319

A:Accession: S09848

A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-586 <CHE>
A:Cross-references: EMBL:X17403; NID:G59591; PIDN:CAA35358.1; PID:G1780862
A:Note: this sequence was submitted to the EMBL Data Library, December 1989
C:Superfamily: human cytomegalovirus 65K early nonstructural protein
C:Keywords: leucine zipper; nonstructural protein
F:114-135/Region: leucine zipper motif
F:324-372/Region: leucine zipper motif

Query Match 6.3%; Score 132; DB 1; Length 586;
Best Local Similarity 23.6%; Pred. No. 1.1;
Matches 104; Conservative 62; Mismatches 145; Indels 130; Gaps 25;
QY 6 RGRGRSR-----GGRRARRARGRCRCRPAROS-----PARLPDPTVLVDL--VSDSDEEV 55
DB 11 RARRPARRGGGGVGSNSRHS-GKRRQRRLSAPPLTFLATTTTMMGVASTDDDS 69
QY 56 LEVADPVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRRLLDGEPVVP 115
DB 70 LLLKTPDEL-----DKYSGSPQITL-----TLTKHDIRQPR 101
QY 116 VYSGK---VOSSNLIPDN-----SSLLKLCPEPEDEADLTNSGSSPSEDALFS 163
DB 102 VHRGTVHLQLHLDRPEELRDPFQILLSTPLQL--GEANDESQ--TAPATLQEEETAAS 157
QY 164 GSPWRKLRKKKEKKMEEPDODISPLQPSRKNRKRKHTALQKLRVKNRQLQDLR 223
DB 158 HEPEKK-----EKQEKKE--DED-----DRNDRER-----GILCVSNESDVR 197
QY 224 SCLS--PKHQSPALQSTDEVLVGGPVLQPSRSLFTKIRCADLVLPLVRMSEPLQN 281
DB 198 PAFSLFPAPRGCHILRSVIDQ-QLTMAIVRLSLNLFALRI-ITPLKRLPLR-----RK 250
QY 282 VVDHMANHGVSPNRILLFGSELSPTATPSTLKLGVDIIDCVVLAASSSATETSQEL 341
DB 251 AAHTALH-----DCIALHLPELTPEPTLDINNVTENAAASVAD---TAESTDAL-DLTPTL 301
QY 342 RLRVQGEKKHQMELISLSPLSPKVLMSHYEAMGLSG-----HKLSD----- 393
DB 302 TVRVR-----HALCWHRVG-GISGRGLTSRISARLSETTAKTLGP 342
QY 384 FFFDGTGLSGKELPADLGLS 404
DB 343 SVFGRLELDPNESPPDLTSS 363

RESULT 9

T09648
nucleolin homolog nuM1 - alfalfa
N:Alternate names: nuM1 protein
C:Species: Medicago sativa (alfalfa)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T09648
R:Boegre, L.; Jonak, C.; Mink, M.; Meskiene, I.; Traas, J.; Ha, D.T.C.; Swoboda, I.; P.
Plant Cell 8, 417-428, 1996
A:Title: Developmental and cell cycle regulation of Alfalfa nucm1 a plant homolog of t
A:Reference number: Z16796; MUID:96361876; PMID:8721748
A:Accession: T09648
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-635 <BOE>
A:Cross-references: EMBL:X88845; NID:G1279562; PIDN:CAA61298.1; PID:G1279563
C:Genetics:
A:Gene: nuM1
C:Superfamily: nucleolin; ribonucleoprotein repeat homology
Query Match 6.1%; Score 128.5; DB 2; Length 635;
Best Local Similarity 20.8%; Pred. No. 1.9;
Matches 92; Conservative 75; Mismatches 183; Indels 93; Gaps 20;
QY 24 ARGRCRPARQSPA-----RLIPDPTVLVDLSD---SDEVLVADPVEVPVARLPAPAK 74


```

      80 AKAPAP-SKPTPAKGNVKKAOPETTSESDDISSDEEVKPAKAVPSKNGSAPVK 138
      75 -----PEODSDSGAAGPAGAPRTLVRRRRRLDPGEAPVVPVSGVQSSINLIP 129
      Db 139 KODESEEDSDSSSDDEKPKPAKAVPSKN-----GSAFAKKDDSEEDSDSDSE 191
      Qy 130 DNSLLKLCPSF-----PEDEADLTNSGSSPSDEDDALPS-----GSPWRKKRKCE 176
      Db 192 DEKPAKAVPSKNGSVSAKKDDSSDEDEDEKPKPAKASKNVSAPTKKAASSSDE 251
      Qy 177 KKKKMEEPPODDISPLQPPSS-RNKRKHTALOKLREVNRLQDLRSCSPKQHQSPA 235
      Db 232 BSDESDDE--DEDAKPVSKPAVAKSKDSSDSDDEDDSSSDDEKPKVASKKEVSESE 309
      Qy 236 LOSTDEVLVEGPVLPOSSRLFTLKIRCADVLRLPVMSBPLOVND-----HMANHL 290
      Db 310 SSSSDDDHKM---NIDKSSDSEDESEDEDEPLKTPQKKITDVE-MYDAGSKGKAPVT 365
      Qy 291 GVSPPRILLPGESLSPATPSTKL-----GVADIIDCVVLASSSE----- 333
      Db 366 PATPN-----ETSGSKTLFVGNLSFDVKRSDIENFGCVGEVVD-VRLASDGGVFKG 417
      Qy 334 -----ATETSGELRLRVQGE-KKQMLEISL-----SPSPKLKLMSHYEAMGLSG 379
      Db 418 FGHVEFAFAEAQAQALEMNGQELHRLALDLARERGAFTPNN-----NSNYSAGSGRG 472
      Qy 380 HKLSFFPDG--TKLSGKELPADL 400
      Db 473 OSQTVFVRGFDKNGLGEDEIRAKL 495

```

RESULT 10

```

B42680
nucleolus-cytoplasm shuttle phosphoprotein - rat
N:Alternate names: Nopp140 protein B
C:Species: Rattus norvegicus (Norway rat)
C>Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text change 17-Nov-2000
C:Accession: B42680; A42680; S27889; S27890; S30510; S30511
R:Meier, U.T.; Blobel, G.
Cell 70, 127-138, 1992
A:Title: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: A42680; MUID:92323542; PMID:1623516
A:Accession: B42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-734 <ME1>
A:Cross-references: GB:M94288
A:Experimental source: clone pTW6
A:Accession: A42680
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 32-180, 'O', 181-734 <ME3>
A:Cross-references: GB:M94287; NID:g205749; PIDN:AAA41718.1; PID:g205750
R:Meier, U.; Blobel, G.
submitted to the EMBL Data Library, May 1992
A:Description: Nopp140 shuttles on tracks between nucleolus and cytoplasm.
A:Reference number: S27889
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-734 <ME2>
A:Cross-references: EMBL:M94288; NID:g205751; PIDN:AAA41719.1; PID:g205752
A:Accession: S27889
A:Molecule type: mRNA
A:Residues: 32-180, 'O', 181-734 <ME12>
A:Cross-references: EMBL:M94287; NID:g205749; PIDN:AAA41718.1; PID:g205750
C:Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C:Keywords: phosphoprotein

```

Query Match 6.1%; Score 128.5; DB 2; Length 734;
 Best Local Similarity 23.2%; Pred. No. 2.3; Indels 95; Gaps 12;
 Matches 71; Conservative 35; Mismatches 105; Indels 95; Gaps 12;

```

      Qy 33 OSPPALIPDTVLVDLSNDE-----VLEVADPVVEVPYARL---PAPAKPEODSDSDSE 84
      Db 347 OSPKAAAQOTOPADSSADSSSESDSSSEEEKTPAKTVSKTPAKPAPVKKAESSSDSS 406
      Qy 85 GAAGBPAGAPRTLVRRRRRLDPGEAPVVPV-----YSGKVOSSLNL 127
      Db 407 DSDSDEDAHPAKPVATSPSLSKPAVTGPKPAKAVATPKOPAGSGQKQSKRABSSSE 466
      Qy 128 IPDNSS-----LTKLCSEPEDEA-----DLTNSGSSPSDED----- 159
      Db 467 BESSSSSEEBATKSVTTTKARVATAKAAPSLPAKQAPRAGGSDSSDSESSSEEEKTPPK 526
      Qy 160 -----ALPSSGSPWRKKLRK-----KCEKEKKMEEPPODDISPLQ--- 195
      Db 527 PPAKKKAAGAAPKPTPVYKKAASESSSSSSSESDSESEEEKK---PKSKATPKQAQKA 582
      Qy 196 ---PSSRN-KSKRHTALOKLREVNKRLQDLRSCSPKQHQSPALOSTDEVLVEGPVL 251
      Db 583 NGVPASQNGKAGKESEEBEDTEONKKAAGTKPGSGKKRKN---ETADDA-----AT 632
      Qy 252 PQSSRL 257
      Db 633 PQSKTV 638

```

RESULT 11

```

D87719
protein R119.4 (imported) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence revision 10-May-2001 #text change 10-May-2001
C:Accession: D87719
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:9905613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A:Accession: D87719
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-742 <STO>
A:Cross-references: GB:chr_I; PIDN:AACT6432.1; PID:g3133012; GSPDB:GN00019; CBSP:R119.4
C:Genetics:
A:Gene: R119.4
A:Map position: 1

```

Query Match 6.1%; Score 128.5; DB 2; Length 742;
 Best Local Similarity 23.0%; Pred. No. 2.3;
 Matches 78; Conservative 33; Mismatches 111; Indels 117; Gaps 14;

```

      Qy 7 GRGPRS-----RGRCARARAGRCPRARQSPARLIPDTVLVDLSNDE--- 52
      Db 170 GAAPSTRGCEKPYVGGGRC-----GARGYSAVAVPSSALEPDATFADIDENQTKVD 224
      Qy 53 BEVLEVADPVE-----VPVAPLPAKPEODSDSGAAGPAGAPRTLVRRRRRL 106
      Db 225 TTVEVQGPVEBSYATVPTTSAPAPL-----SFAVAAAHAHKEARKKQAQNP 274
      Qy 107 DGEAPVVPVYSGKVOSSLNLIPDNSSLLKLCPSPEDEADLTNSGSSPSDEDDALPSSGP 166
      Db 275 QPAAPP-----RRSLSPQPP-----LPSTVAP 295
      Qy 167 WRKK-----LRKKCEKKEKKMEFPPODDISPLQPPSSRNKRKHTALOKLREVNRL 219
      Db 296 VKEEPAAPVFPPEPTSTQPEKEDFPQNESSVLAEEQTPNVSTHDEVQS----- 347
      Qy 220 QDLRSLSPKQHQSPALO-STDEVLVEGPVLPOSSRLFTLKIRCADVLRLPVMSBP 278
      Db 348 -----TPBPQAWTQTKTDLIGLSEAPQLGIS-----PIPSAP 383
      Qy 279 LQNVVDHMANHLG-VSPNRL-LRGESELSP-ATPST 314
      Db 384 VQIIPDPGVEFVGTTAPTNIDYSGFVEAAPSQLPST 422

```



```

QY 101 RRRLLD---PGAEVVPVYSGVQSSLNLLIPDNSS---LLKLCPSPEDEADLTNGS 153
Db 483 PTKKPADPKPSEBDEKVPAPKMKPMEEDDDDEPADFTVPIRGEDEDEPADDBE 542
QY 154 SPSEDDALPSSGPMKRIKCKCEKEKKEEFPDODISLPDP----- 196
Db 543 EPEDE---PABDEPKKKPKKRRKKRKKKPPVEPEKEPTPEPVVPAKPIAKKEBE 599
QY 197 -----SSRYSKRKHTALOKLREVNKRLQDLRSCLSPKOHOSPAL----- 236
Db 600 PIPMPKPKTIARERKERIPALRYAKK-PRELEVYIPFVIMW-QTAALLTQEGMGAF 657
QY 237 ---OSTDDEVVLVEGPVL---QSSRLFTL-----KIRCRADLVRLPYRMSEPLQNVVDH 285
Db 658 GKSRAANEVNFGRPIVQGAVDSTKVIPLMNDSEKCANRSGMTAFGA-PREIDQNVVDH 716
QY 286 MANHGVSPNRILLFGSELSPTATPSTLKGVADI 322
Db 717 --HKGKSQGIPLAKGTYHPHGEYGTIRROTADV 750

```

RESULT 15

```

JM0057
Gravlin - human
C/Species: Homo sapiens (man)
C/Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C/Accession: JMW0057
R/Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A/Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial
A/Reference number: JMW0057; MUID:98269042; PMID:9604001
A/Accession: JMW0057
A/Molecule type: mRNA
A/Residues: 1-1684 <SAT>
A/Cross-references: DDBJ:AB003476; NID:G2081606; PIDN:BA19927.1; PID:d1020716; PID:G208
C/Comment: This protein regulates cell growth.
F/433-439/Region: nuclear location signal
F/522-527/Region: nuclear location signal
F/591-596/Region: nuclear location signal
F/671-676/Region: nuclear location signal

```

Query Match 5.9%; Score 124.5; DB 2; Length 1684;
 Best Local Similarity 20.7%; Pred. No. 11;
 Matches 94; Conservative 70; Mismatches 171; Indels 119; Gaps 21;

```

QY 43 VLVVLSDSDEVLVADPVVPAVRLPAPAKPEODSDS----- 82
Db 32 VVHDTDDGOETPEIIOIPSESNLELTQPTESQANDIGKKVFKVGFVKTKDK 91
QY 83 -----SEGAAEGPAGA-----PRLVRRRRRLDPGEAPVVPVYSGK-VQSS 124
Db 92 TEKPDVOLLTVKDDGEGACAGCDKPSLGAAGAASEPKOSTEKPEETLKREOSH 151
QY 125 LNLIP--DNSSLKLCPSPEPD--EADLTNGSSPSEDDALPSSGSP-----WRK 169
Db 152 AEISPPAESQAVEBEKEGEKEKESKSAESPTSPVTSSTGTFKKFTQGWAGWRK 211
QY 170 K--LRKKKEK-----EKMEEFPPDODISPLPPSSRNKSRKHTALQKLRVNRKLDLR 223
Db 212 KTSFPKPEDEVAEAKKEKEPEK-----VTDEEDGAEVAEKEKLTASSEQAHPQ-EPAE 265
QY 224 SCLSPK--OHSPALOSTDDEVVLVEGP---VLPSSRLFTLKIRCRADLVRLPYRMS 276
Db 266 SAHERLSAEYKVELPS-EEVSGSQSPSEKPRAPLATEVDEKIEVHQEEVVAEYHVS 324
QY 277 EPLQNVDMANHLGVSPNRILLFGSELSPTATPSTLKGVADIIDCVVLASSSEATE 336
Db 325 TVERTEBO-----KTEVERETA-----GSVPABELVEMDAEPQEA 360
QY 337 TSQEL-RLR---VQCKEKGMLEISLSPDSPL-----KVLMSHYEAMGLSGHKL 382
Db 361 PAKELVKKETCVSGEDPTQAD--LSPDEKVLSPREGVSEVEMLSQERMKVQSGSPL 418

```

```

QY 383 SFFPDGT---KLSGKEPLADLG---LESGLIEV 410
Db 419 KKLFTSTGLKKLSGKKQKQKRGGBESGERTQV 452

```

Search completed: January 12, 2003, 10:28:03
 Job time : 28 secs

This Page Blank (uspc)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 10:23:09 ; Search time 14 Seconds

(without alignments)
1220.589 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099
Sequence: 1 MAEPPLRGGRGSRGGRGARR.....GKELPADLGLSGDLIEWWG 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	135.5	6.5	831	1 NPH_RAT	P16884 ratnus norv
2	134.5	6.4	587	1 UH84_HCMWT	P28839 human cytom
3	132	6.3	586	1 UH84_HCMVA	P16727 human cytom
4	128.5	6.1	704	1 NP14_RAT	P41777 ratnus norv
5	128	6.1	303	1 VG08_BP22	P26748 bacterioph
6	127	6.1	1365	1 SU22_DROME	P25172 drosophila
7	122.5	5.8	1781	1 AKAC_HUMAN	Q02952 homo sapien
8	121	5.8	1109	1 TCF8_RAT	Q62947 ratnus norv
9	119	5.7	1603	1 PSC_DROME	Q35820 drosophila
10	118.5	5.6	718	1 TRF2_CHICK	Q95653 gallus gall
11	116.5	5.6	1039	1 MSU1_DROME	P34739 drosophila
12	116.5	5.6	1061	1 LUS_DROME	Q60542 mesocricetu
13	116	5.5	1043	1 TCF8_MESAU	Q34758 saccharomyc
14	115.5	5.5	872	1 SCDS_YEAST	Q13428 homo sapien
15	114.5	5.5	1411	1 TCOF_HUMAN	P28023 ratnus norv
16	113.5	5.4	1280	1 DYNA_RAT	P18858 homo sapien
17	112.5	5.4	95	1 SM32_HUMAN	Q01188 homo sapien
18	112	5.3	919	1 DNLI_HUMAN	Q01188 homo sapien
19	112	5.3	943	1 CENC_HUMAN	Q01188 homo sapien
20	112	5.3	1004	1 PRPX_RAT	Q64547 ratnus norv
21	111.5	5.3	1170	1 XPG_MOUSE	P35689 mus musculu
22	111	5.3	741	1 BSG2_DROME	P11929 drosophila
23	111	5.3	897	1 EP15_MOUSE	P45567 mus musculu
24	111	5.3	1163	1 RTN4_RAT	Q91111 ratnus norv
25	111	5.3	1281	1 DYNA_MOUSE	Q08788 mus musculu
26	110.5	5.3	854	1 CLPB_THETH	Q94633 thermus the
27	110.5	5.3	919	1 SYNP_HUMAN	Q94618 h nuclear r
28	110.5	5.3	2517	1 NCR2_HUMAN	Q94618 h nuclear r
29	110	5.2	542	1 TULI_HUMAN	Q00294 homo sapien
30	110	5.2	612	1 ARRS_MAIZE	P13027 zea mays (m
31	110	5.2	883	1 E74B_DROME	P11536 drosophila
32	110	5.2	1206	1 FM14_MOUSE	Q05859 mus musculu
33	110	5.2	2150	1 SDC3_CAEEL	P34706 caenorhabdi

ALIGNMENTS

RESULT 1	ID	NFH_RAT	STANDARD	PRT	831 AA.
AC	P16884	Q63368			
DT	01-AUG-1990	(Rel. 15, Created)			
DT	01-FEB-1994	(Rel. 28, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Neurofilament triplet H protein (200 kDa neurofilament protein)				
DE	(Neurofilament heavy polypeptide) (NF-H) (Fragment).				
GN	NFPH OR NFH.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_TaxID=10116;				
OX	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RX	MEDLINE=89065087; PubMed=3143606;				
RA	Breen K.C., Robinson P.A., Wion D., Anderson B.H.;				
RT	"Partial sequence of the rat heavy neurofilament polypeptide (NF-H).				
RT	Identification of putative phosphorylation sites."				
RT	FEBS Lett. 241:213-218(1988).				
RM	[2]				
RP	SEQUENCE OF 37-831 FROM N.A.				
RX	MEDLINE=88309090; PubMed=2457365;				
RA	Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L.,				
RA	Jolles P.;				
RT	"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and				
RT	in situ detection."				
RL	Biochem. Biophys. Res. Commun. 154:1099-1106(1988).				
RM	[3]				
RP	SEQUENCE OF 1-89 AND 243-313 FROM N.A.				
RX	MEDLINE=87080760; PubMed=2878828;				
RA	Robinson P.A., Wion D., Anderson B.H.;				
RT	"Isolation of a cDNA for the rat heavy neurofilament polypeptide				
RT	(NF-H)."				
RL	FEBS Lett. 209:203-205(1986).				
RM	[4]				
RP	SEQUENCE OF 318-831 FROM N.A.				
RX	MEDLINE=89184647; PubMed=2928342;				
RA	Liubertow I., Spinner N., Snyder S., Anderson J., Goldsaber D.,				
RA	Smolowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;				
RT	"Cloning of a cDNA encoding the rat high molecular weight				
RT	neurofilament peptide (NF-H): developmental and tissue expression in				
RT	the rat, and mapping of its human homologue to chromosomes 1 and				
RT	22."				
RU	Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).				
CC	-I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,				
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.				
CC	NF-H HAS AN IMPORTANT FUNCTION IN NATURE AXONS THAT IS NOT				
CC	SUBSERVED BY THE TWO SMALLER NF PROTEINS.				
CC	-I- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS				
CC	PHOSPHORYLATED ON A NUMBER OF THE SERIES IN THIS MOTIF. IT IS				
CC	THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF				
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE				
CC	OF AXONAL CALIBER.				


```

Db 257 ALHDCMAHL-----PELTRESTLDINNNTENASVADAESTDADLT 299
Qy 339 QELRLRVOGKEKQMLLEISLSPDPLKY-----LMSHYEAMGLSGKLSFFPDGTLKSG 393
Db 300 PLTVRVHVAOCMHVEGGISGPRGLTSRISARLSETTAKTLGSP-----VFGRLLEIDP 353
Qy 394 KFLPADLGLES 404
Db 354 NESPDLTLLSS 364

RESULT 3
UL84_HCMVA
ID UL84_HCMVA STANDARD; PRT; 586 AA.
AC P16727;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE 65 kDa early nonstructural protein (UL84 protein).
GN UL84.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10360;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horsnell T., Hutchinson S.C., Ili, Kouzarides T., Martignetti J.A.,
RA Peddie E., Satchwell S.A., Tomlinson P., Weston K.M., Barrell B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X17403; CAAS3558.1; -.
DR PIR; S09848; WMBEDE.
KM Nonstructural protein.
FT DOMAIN 9 19 ARG-RICH (BASIC).
FT DOMAIN 162 170 LYS-RICH (BASIC).
FT DOMAIN 171 182 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 586 AA; 65428 MW; 54AB912D6077223F CRC64;

Query March 6.3%; Score 132; DB 1; Length 586;
Best Local Similarity 23.6%; Pred. No. 0.47; Indels 130; Gaps 25;
Matches 104; Conservative 62; Mismatches 145;

Qy 6 RGRGRSR-----GGGARARAGARCRPAROS-----PARLPTVLVDL--VSDSDEV 55
Db 11 RARPRARAGGGGAGVSSSRHS-GKCRORRALAPLTLFATTTTMMGVASTDDDS 69
Qy 56 LEVADPVVPAARLPAPAKPEODSDSSEGAEGPAGPRTLVRRRRLDPCGAPVP 115
Db 70 LLKTPDEL-----DKYSGSPQTL-----TLTDKHIROP 101
Qy 116 VYSGK---VOSSLNLPN-----SSLKLCPSPEPEADLTNGSSPEDDALPS 163
Db 102 VHRGTHIQLHLDRPELDPPQILSTPQL--GEANDESQ--TAPATIQEETPAAS 157
Qy 164 GSPWRKRLRKCKEKEKKEEPPDDISPLPOPSRRKSKTEALQKLRVKNKRLQDLR 223
Db 158 HEPEKKK-----EKQEKKE--DED-----DRNDRER-----GLTCVSNEDSDVR 197
Qy 224 SCLG--PKOHOSPALQSDDEVDVVEGVPVLPOSSRLFLTKRCRADVRLPRMSEPIQN 281
Db 198 PAFSLFPAFGCHILRSVIDQ-QLTRMAIVRLSLNLFALRI-ITPLKRLPLR-----RK 250

```

```

Qy 282 VVDHANHLGVSFNRIILLFSESELSPATPSTLKGVAIDIIDCVLLASSSEATSTOEL 341
Db 251 AAHHTALH-----DCLALHLPELTFPELTLDINNNTENASVAD---TAESTD A-DLTPRL 301
Qy 342 RLRLVOGKEKQMLLEISLSPDPLKYLMHYEAMGLSG-----HKL----- 383
Db 302 TVRV-----NALCWHRVEG--GISGPRGLTSRISARLSETTAKTLGSP 342
Qy 384 FFPDGTKLSGKELPADLGLES 404
Db 343 SVFGRLEIDPNESPDLTLLSS 363

RESULT 4
NP14_RAT
ID NP14_RAT STANDARD; PRT; 704 AA.
AC P41777;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE nucleolar phosphoprotein) (Nop140) (Nucleolar and coiled-body
DE phosphoprotein 1).
GN NOLC1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RX SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.
RX MEDLINE=92323542; PubMed=1623516;
RA Meier U.T., Blobel G.;
RT TISSUE=Liver;
RT "Nop140 shuttles on tracks between nucleolus and cytoplasm."
RL Cell 70:127-138(1992).
CC [2]
CC INTERACTION WITH NOP5 AND FIBRILLARIN.
CC MEDLINE=20143579; PubMed=10679015;
CC Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;
CC "Conserved composition of mammalian box H/ACA and box C/D small
CC nucleolar ribonucleoprotein particles and their interaction with the
CC common factor Nop140."
CC Mol. Biol. Cell 11:567-577(2000).
CC -1- FUNCTION: RELATED TO NUCLEOGENESIS. MAY PLAY A ROLE IN THE
CC MAINTENANCE OF THE FUNDAMENTAL STRUCTURE OF THE FIBRILLAR CENTER
CC AND DENSE FIBRILLAR COMPONENT IN THE NUCLEOLUS. IT HAS INTRINSIC
CC GTPASE AND ATPASE ACTIVITIES. MAY PLAY AN IMPORTANT ROLE IN
CC TRANSCRIPTION CATALYZED BY RNA POLYMERASE I (BY SIMILARITY).
CC -1- SUBUNIT: Interacts with DKC1/Nap57, NOP5/Nap65 and fibrillarin.
CC -1- SUBCELLULAR LOCATION: SHUTTLES ON CURVILINEAR TRACKS BETWEEN
CC NUCLEOLUS AND CYTOPLASM. THESE TRACKS EXTEND FROM THE DENSE
CC FIBRILLAR COMPONENT OF THE NUCLEOLUS ACROSS THE NUCLEOLAR PLASM TO
CC A LIMITED NUMBER OF NUCLEAR PORE COMPLEXES.
CC -1- PTM: THIS PROTEIN UNDERGOES RAPID AND MASSIVE PHOSPHORYLATION AND
CC DEPHOSPHORYLATION ON CK-II AND PKC SITES. NOP140 IS ONE OF THE
CC MOST PHOSPHORYLATED PROTEINS IN THE CELL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M94287; AAA41718.1; -.
DR EMBL; M94288; AAA41719.1; -.
KM Nucleolar protein; phosphorylation; Repeat; GTP-binding; ATP-binding.
FT DOMAIN 84 570
FT REPEAT 84 95 ACIDIC SERINE CLUSTER.
FT REPEAT 127 138 ACIDIC SERINE CLUSTER 2.

```



```

Db 799 NSVPC-----LRALANKQTLLIPQVATYSATVSPAMQEPVAVIQP 841
Qy 369 -SHYEFAMGLSGHKLFSF-----FFDGTGLSGKELPADLIESGLIE 409
Db 842 NQNDQROPTSSSEGVSVEDQNDSDCTPPKKTKRKAMGNVACDLCD 887

RESULT 9
PSC_DROME
ID_PSC_DROME STANDARD: PRT: 1603 AA.
AC P35820:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Posterior sex combs protein.
GN PSC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RA "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
RT encode proteins with homology to the murine bml-1 oncogene.";
RL Nature 353:351-353(1991).
CC -1- FUNCTION: THE POLYCOMB GROUP (PC-G) GENES ARE NEEDED TO MAINTAIN
CC EXPRESSION PATTERNS OF THE HOMEOBOX SELECTOR GENES OF THE
CC ANTENNAPEDIA (ANTP-C) AND BITHORAX (BX-C) COMPLEXES, AND HENCE FOR
CC THE MAINTENANCE OF SEGMENTAL DETERMINATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59375; CAA41965.1; -
DR PIR; S17983; S17983.
DR FlyBase; FBgn0005624; Bsc.
DR InterPro; IPR001841; Znf_fing.
DR Pfam; PF00097; ZF-C3HC4.1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT DOMAIN 47 53 POLY-THR.
FT DOMAIN 83 88 POLY-THR.
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 152 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT ZN_FING 265 304 RING-TYPE.
FT DOMAIN 642 651 POLY-SER.
FT DOMAIN 1066 1069 POLY-GLY.
FT DOMAIN 1185 1189 POLY-PRO.
FT DOMAIN 1214 1217 POLY-PRO.
FT DOMAIN 1391 1396 POLY-PRO.
FT DOMAIN 1458 1461 POLY-ALA.
FT DOMAIN 1517 1520 POLY-GLY.
SQ SEQUENCE 1603 AA; 169999 MW; 77024F4097736473 CRC64;
Query Match 5.7%; Score 119; DB 1; Length 1603;
Best Local Similarity 18.9%; Pred. 9.8;
Matches 68; Conservative 54; Mismatches 123; Indels 114; Gaps 13;
Qy 45 VDLVSDSBEVLEVDAPVEVVARLLPAPAKPEQSDSDS---DSEGAAGPAGAPRTLVRRR 101

```

```

Db 598 IDLSKONSVTIIDMSDPERREIVK---PLKPEKESRKXKDXGSPKSSSSSSSSGGR 654
Qy 102 RRLRLDGPAPVAVPYSGVOSS---LNLIP-----DNSLLKLCPSRP----- 142
Db 655 KKKSPSLTPVPLTRTERINSPSGVSTLSPRYTSGATSEDPKSEFLKPKLVKYE 714
Qy 143 EDEADLTNNGSSSPSEDDALPSGSPWRKKLRKKCEKKEEFPQDQISPLPQPSRRNS 202
Db 715 SPERTLNRRATPPSPVQGSASP-----KSKGNLNDOSILMKKPPSCMPKSIASKRS 769
Qy 203 RKTALQKLRVKNRKLQDLRSCLSPPKHQSPALOSTDEVVL-----VEG 248
Db 770 KEPVAVSK-----KQKLSPLPTVPFKIRLPTVNGSSGTSAPKIEK 812
Qy 249 PVLQSSR---LFTLKIRCRADVLRLP-----VRMSEP-----LQNVVDHMA- 287
Db 813 PLMPPPAKPMPLAPRKLQPSQAPAPPSPFIHHAGVMSAPGNRTPIAKRYQITLPKASR 872
Qy 288 -NHLGVSPNRILLP-----GESELSPTATPST 314
Db 873 PVFPANIPDVVRLKADGTEIKSIGGGSVENNSNAQKPHLYGPKGTVMGPALPAT 931

RESULT 10
TRF2_CHICK
ID_TRF2_CHICK STANDARD: PRT: 718 AA.
AC Q9PU53:
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Telomeric repeat binding factor 2 (TTAGCG repeat binding factor 2)
DE (Telomeric DNA binding protein).
GN TRF2 OR TRF2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.; AND VARIANT VAL-438.
RC STRAIN=White Leghorn; TISSUE=Embryo;
RX MEDLINE=20035814; PubMed=10571037;
RA Konrad J.P., Mills W., Easty D.J., Farr C.J.;
RA "Cloning and characterisation of the chicken gene encoding the
RT telomeric protein TRF2.";
RL Gene 239:81-90(1999).
CC -1- FUNCTION: Binds the telomeric double-stranded TTAGCG repeat.
CC Protects against end-to-end fusion of chromosomes and plays a role
CC in successful progression through the cell division cycle (By
CC similarity).
CC -1- SUBUNIT: Homodimer. Binds to RAP1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear. Colocalizes with telomeric DNA in
CC interphase cells and is located at chromosome ends during
CC metaphase (By similarity).
CC -1- TISSUE SPECIFICITY: Highly expressed in embryo.
CC -1- SIMILARITY: CONTAINS 1 MYB-LIKE DOMAIN.
CC -1- CAUTION: The sequence was modified by adding the 2 missing N-
CC terminal residues based on the sequence of the genomic clone.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ133783; CAB56220.1; -
DR HSSP; P54774; 18A5.
DR InterPro; IPR001005; Myb_DNA_binding.
DR Pfam; PF00249; myb_DNA_binding; 1.
DR SMART; SM00395; SANT; 1.

```



```

Oy 372 BEAMGLSGH 380
Db 703 KEELISQSH 711

RESULT 12
IDS_DROME STANDARD; PRT; 1061 AA.
ID IDS_DROME
AC P34739;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable helicase lodestar.
GN Lds.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OC NCB1_TaxID=7227;
RN (1)
RX MEDLINE=92009170; PubMed=1916263;
RA Girdham C.G., Glover D.M.;
RT "Chromosome tangling and breakage at anaphase result from mutations
RT in lodestar, a Drosophila gene encoding a putative nucleoside
RT triphosphate-binding protein.";
RL Genes Dev. 5:1786-1799(1991).
RN [2]
RN CONCEPTUAL TRANSLATION.
RX MEDLINE=93181281; PubMed=8382805;
RA Bork P., Koonin E.V.;
RT "An expanding family of helicases within the 'DEAD/H' superfamily.";
RL Nucleic Acids Res. 21:751-752(1993).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
CC BE INTRODUCED IN POSITION 946 TO PRODUCE THIS ORF.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; X62629; CA444496.1; ALT_FRAME.
DR PIR; A40580; A40580.
DR Flybase; FBgn002542; lds.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00176; SNF2_N; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC_C; 1.
DR Nucleic protein; Helicase; ATP-binding.
KW NP_BIND
FT SITE 465 472 ATP (POTENTIAL).
FT SITE 603 606 DEAD BOX.
SQ SEQUENCE 1061 AA; 118189 MW; E82AA64254342B80 CRC64;

Query Match 5.6%; Score 116.5; DB 1; Length 1061;
Best Local Similarity 22.2%; Pred. No. 8.4;
Matches 92; Conservative 50; Mismatches 144; Indels 129; Gaps 20;

```

```

Oy 130 DNSLLKLCPSPEDEADLTNSGSSPSSEDDALP-----SGSPWRKKLR----- 173
Db 194 DSSGSDVLLLSNKETPIELSS-----TDDDATNNKNNSGPFERPSXLSPPSSAGASV 249
Oy 174 -KCEK--BEKKKEFPDDDISPLPPSSRANKSR-----HTALQKRPVNRLODL 222
Db 250 VRTSKNLSQPTTQAVLKQKTPAPRRSRIRKSEDKVVSQVYVDEMKRLAEKRVQSDA 309
Oy 223 RSCLSPKQHSPLAQSTDEVVLVEGVLPOSSRLFTLKIRCADVLVLPVMSPELQNV 282
Db 310 EKLFEKVAHKLPDKKS-----QIMKRIDTLARELAMDE--QMI 345
Oy 283 VDHMANHLGVSPNRIILLFGESSELPATPSTLKGVDIIDCVVLASSSEATETSGELR 342
Db 346 SALRVQGSNVPAVRV-----KPTLNPRAPS-----IDTLMDDELSEAVNEIKPVY 392
Oy 343 LRVQ-----KEKHQMLEISLP-----DSP--LKY-LMSHYEAM 375
Db 393 TGAQGMATFNQKALTLESXLDLHVSLDLPQPEVLAEPPGLKVLNMHQHAL 447

RESULT 13
TCF8_MESAU STANDARD; PRT; 1043 AA.
ID TCF8_MESAU
AC Q60542;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor 8 (Zinc finger protein BZF).
GN TCF8 OR BZF.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCB1_TaxID=10036;
RN (1)
RN CONCEPTUAL TRANSLATION.
RX MEDLINE=95021206; PubMed=7935395;
RA Franklin A.J., Jettion T.L., Shelton K.D., Magnuson M.A.;
RT "BZF, a novel serum-responsive zinc finger protein that inhibits gene
RT transcription.";
RL Mol. Cell. Biol. 14:6773-6788(1994).
CC -!- FUNCTION: ACTS AS A TRANSCRIPTIONAL REPRESSOR. BINDS TO E-BOX
CC SEQUENCES IN THE IMMUGLOBULIN HEAVY CHAIN ENHANCER AS WELL AS IN
CC THE REGULATORY REGIONS OF MANY OTHER TISSUE-SPECIFIC GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO DELTA-EF1/ZFH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMEDOMAIN PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC -----
DR EMBL; U13856; AAC37667.1; -.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2; 7.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 2.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; zinc-finger;
KW Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT ZN_FING 94 117 C2H2-TYPE.
FT ZN_FING 124 146 C2H2-TYPE.

```


DB 467 IK-----P5STPETYNSGGLQPLKPTATGSANYLMKHQHSVUNNPV 509

RESULT 15

TCOF HUMAN STANDARD: PRT: 1411 AA.

ID TCOF HUMAN 013428: 099408: 099860: 01-NOV-1997 (Rel. 35, Last sequence update)

AC 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

BT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Treacle protein (Treacher collins syndrome protein).

GN TCOF1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96154183; PubMed=8563749;

RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K., Bonner C.A., Koprivnikar K., Wasmuth J.U.;

RT "Positional cloning of a gene involved in the pathogenesis of Treacher Collins syndrome";

RU Nat. Genet. 12:130-136(1996).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97228900; PubMed=9074926;

RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J., Dixon M.J.;

RT "Identification of the complete coding sequence and genomic organization of the Treacher Collins syndrome gene";

RU Genome Res. 7:223-234(1997).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=97250498; PubMed=9096354;

RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M., Ashley J.A., Lovett M., Jabs E.W.;

RT "TCOF1 gene encodes a putative nuclear phosphoprotein that exhibits mutations in Treacher Collins syndrome throughout its coding region";

RU Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).

RN [4]

RP VARIANTS L-439; V-810; V-1113 AND G-1355, AND VARIANT TCS R-53.

RX MEDLINE=97195537; PubMed=9042910;

RA Edwards S.J., Gladwin A.J., Dixon M.J.;

RT "The mutational spectrum in Treacher Collins syndrome reveals a predominance of mutations that create a premature-termination codon";

RU Am. J. Hum. Genet. 60:515-524(1997).

CC -1- DISEASE: DEFECTS IN TCOF1 ARE THE CAUSE OF TREACHER COLLINS SYNDROME (TCS). TCS IS A AUTOSOMAL DOMINANT DISORDER OF CRANIOFACIAL DEVELOPMENT THAT OCCURS WITH AN INCIDENCE OF 1/50,000 LIVE BIRTHS. THE CLINICAL FEATURES OF TCS ARE BILATERALITY, SYMMETRICAL AND INCLUDE: (1) ABNORMALITIES OF THE EXTERNAL EARS, ARESTIA OF THE EXTERNAL EAR CANALS, AND MALFORMATION OF THE MIDDLE EAR OSSICLES, WHICH MAY RESULT IN CONDUCTIVE HEARING LOSS; (2) LATERAL DOWARD SLOPING OF PALPEBRAL FISSURES, FREQUENTLY WITH COLOBOMAS OF THE LOWER EYELIDS; (3) HYPOPLASIA OF THE MANDIBLE AND ZYGOMATIC COMPLEX; (4) CLEFT PALATE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U40847; AAC50903.1; -

DR EMBL; U76366; AAC51181.1; -

DR EMBL; U84644; AAC51185.1; -

DR EMBL; U84640; AAC51185.1; JOINED.

DR EMBL; U84641; AAC51185.1; JOINED.

DR EMBL; U84642; AAC51185.1; JOINED.

DR EMBL; U84643; AAC51185.1; JOINED.

DR EMBL; U84644; AAC51185.1; JOINED.

DR EMBL; U84645; AAC51185.1; JOINED.

DR EMBL; U84646; AAC51185.1; JOINED.

DR EMBL; U84647; AAC51185.1; JOINED.

DR EMBL; U84648; AAC51185.1; JOINED.

DR EMBL; U84649; AAC51185.1; JOINED.

DR EMBL; U84650; AAC51185.1; JOINED.

DR EMBL; U84651; AAC51185.1; JOINED.

DR EMBL; U84652; AAC51185.1; JOINED.

DR EMBL; U84653; AAC51185.1; JOINED.

DR EMBL; U84654; AAC51185.1; JOINED.

DR EMBL; U84655; AAC51185.1; JOINED.

DR EMBL; U84656; AAC51185.1; JOINED.

DR EMBL; U84657; AAC51185.1; JOINED.

DR EMBL; U84658; AAC51185.1; JOINED.

DR EMBL; U84659; AAC51185.1; JOINED.

DR EMBL; U84660; AAC51185.1; JOINED.

DR EMBL; U84661; AAC51185.1; JOINED.

DR EMBL; U84662; AAC51185.1; JOINED.

DR EMBL; U84663; AAC51185.1; JOINED.

DR EMBL; U79659; AAB40722.1; JOINED.

DR EMBL; U79645; AAB40722.1; JOINED.

DR EMBL; U79646; AAB40722.1; JOINED.

DR EMBL; U79647; AAB40722.1; JOINED.

DR EMBL; U79648; AAB40722.1; JOINED.

DR EMBL; U79649; AAB40722.1; JOINED.

DR EMBL; U79650; AAB40722.1; JOINED.

DR EMBL; U79651; AAB40722.1; JOINED.

DR EMBL; U79652; AAB40722.1; JOINED.

DR EMBL; U79653; AAB40722.1; JOINED.

DR EMBL; U79654; AAB40722.1; JOINED.

DR EMBL; U79655; AAB40722.1; JOINED.

DR EMBL; U79656; AAB40722.1; JOINED.

DR EMBL; U79657; AAB40722.1; JOINED.

DR EMBL; U79658; AAB40722.1; JOINED.

DR EMBL; U79659; AAB40722.1; JOINED.

DR MIM; 154500; -

DR InterPro; IPR003993; treacle.

DR Pfam; PF03546; treacle; 3.

DR PRINTS; PRO1503; TREACLE.

KW Disease mutation; Polymorphism.

FT DOMAIN 89 97 POLY-GU.

FT DOMAIN 204 207 POLY-SER.

FT DOMAIN 616 619 POLY-SER.

FT DOMAIN 919 924 POLY-SER.

FT DOMAIN 1285 1289 POLY-LYS.

FT DOMAIN 1375 1386 POLY-LYS.

FT DOMAIN 1398 1405 POLY-LYS.

FT VARIANT 53 53 W -> R (IN TCS).

FT VARIANT 439 439 /FTID=VAR_005630.

FT VARIANT 439 439 P -> L.

FT VARIANT 810 810 /FTID=VAR_005631.

FT VARIANT 810 810 A -> V.

FT VARIANT 1313 1313 /FTID=VAR_005632.

FT VARIANT 1313 1313 A -> V.

FT VARIANT 1355 1355 /FTID=VAR_005633.

FT VARIANT 1355 1355 D -> G.

FT CONFLICT 1312 1312 /FTID=VAR_005634.

FT SEQUENCE 1411 AA; 144312 MW; 38802030D985C2699 CRC64; K -> O (IN REF. 2).

Query Match 5.8%; Score 114.5; DB 1; Length 1411;

Best Local Similarity 21.4%; Pred. No. 16;

Matches 84; Conservative 50; Mismatches 172; Indels 87; Gaps 16;

OY 2 AEPPLRGGRPRSG-----GCGARARGARCRPRAROSPARTIDTVLVDSDEEVL 56

DB 715 ANPAAARAPSAKGTISAPGKVTAAQAQAKORSKVPV-----NPNSTV 762

OY 57 EVADPEVVP-VARLPAPAK-----PEQDS-----DSDEGAEGPAGAPRTLVRRRRRL 105

Db 763 LARGPASVPSVGKAVATAAAQAQTGPBEDSGSSEESDEEAEETLAAQKPSGKTHOIRAA 822
Qy 106 LDP-----GEAPVVPVYSGKVQSSNLNLPDNNSSLLKLCPSSEPEDEADLTNSG----- 152
Db 823 LAFAKESPRKGAAPTTPGKTGPAAQAQKODDSGSSSEESDSGGEAPAAVTSQAQVIKPPL 882
Qy 153 --SSPSEDDALPSGSPWR-----KKLRKKCKEKKMEFFPDQDISPLPQ---PSSRN 200
Db 883 IFVDPNRSFAGPAATPAQAQAASTPRKARASESTARSSSESEDEVDVIPATQCLTPGIRT 942
Qy 201 K-----SRKTEALQKLEVNKRLQDLRLSCLSPKHQSPALQSTDDDEVVLVEGP 249
Db 943 NVVTMTAHPRIAPKASMAGASSKSSRISD-----GKKQEGPATQ-----VSKKNP 990
Qy 250 V-LPQSSRLFTLKIRC-RADLVRLPVRMSEPLQNVVDHMANHLGVSPNRIILLFGESELS 307
Db 991 ASLPLTQA--ALKVLAQKASEAQPPVARTQPSGSGVDSAVGTLPATSPQ-----STSVQ 1041
Qy 308 PTATPSTLKLGVADIIDCVVLASSSEATETSQE 340
Db 1042 AKGTNKLKRPKLPVQQAATKAPESDDSDSDSD 1074

Search completed: January 12, 2003, 10:26:53
Job time : 20 secs


```
QY 121 VQSSNLIPDNSSLLKLCPEPEDEADLTNSGSSPSSEDDALPSGSPWRKKLRKKCKEKEK 180
Db 121 VQSSNLIPDNSSLLKLCPEPEDEADLTNSGSSPSSEDDALPSGSPWRKKLRKKCKEKEK 180
QY 181 KMEEFPPQDQISPLPQPSSRNKSRKHTEALQKLRVKNRQLDLRSCLSPKQHQSPALQST 240
Db 181 KMEEFPPQDQISPLPQPSSRNKSRKHTEALQKLRVKNRQLDLRSCLSPKQHQSPALQST 240
QY 241 DEVVLVEGPVLPQSSRLFTLKIRCRADLRLPVRMSEPLQNVVDHMANHLGVSPNRILL 300
Db 241 DEVVLVEGPVLPQSSRLFTLKIRCRADLRLPVRMSEPLQNVVDHMANHLGVSPNRILL 300
QY 301 FGSELSPTATPTTLKLGVAIDICVVLASSSEATETSQELRLRVQKQKHOMLEISLSP 360
Db 301 FGSELSPTATPTTLKLGVAIDICVVLASSSEATETSQELRLRVQKQKHOMLEISLSP 360
QY 361 DSPKVLMSHYEAMGLSGHKLSPFFDGTGTLGSKELPADLGESGDLIEVWG 412
Db 361 DSPKVLMSHYEAMGLSGHKLSPFFDGTGTLGSKELPADLGESGDLIEVWG 412

RESULT 2
Q9CVY5 PRELIMINARY; PRT; 416 AA.
AC Q9CVY5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Nuclear factor of activated T-cells, cytoplasmic 2 interacting protein
DE (Fragment).
GN NFAIC2IP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotojori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya I., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK005947; BAB24331.1; -.
DR MGD; MGI:1329015; Nfatc2ip.
DR InterPro; IPR000626; Ubiquitin.
DR SMART; SM00213; UBQ; 1.
FT NON TER 1.
SQ SEQUENCE 416 AA; 45551 MW; E4B46F65CC571AF5 CRC64;
```

Query Match 94.4%; Score 1982; DB 11; Length 416;
Best Local Similarity 97.0%; Pred. No. 4.3e-132;
Matches 394; Conservative 1; Mismatches 9; Indels 2; Gaps 1;

```
QY 9 GPSSRGCGARRA--RGARGCPRARQSPARLIPDTVLVDVSDSEEVLEVPV 66
Db 11 GTWSEVPWRPRRSETRGARGCPRARQSPARLIPDTVLVDVSDSEEVLEVPV 70
```

```
QY 67 ARLPAPAPQEQSDSDSGAAGGPAGAPRTLVRRRRRLLDPGEAPVVPVYSGKVQSSLN 126
Db 71 ARLPAPAPQEQSDSDSGAAGGPAGAPRTLVRRRRRLLDPGEAPVVPVYSGKVQSSLN 130
QY 127 LIPDNSSLLKLCPSPEDEADLTNSGSSPSSEDDALPSGSPWRKKLRKKCKEKEKMEBFP 186
Db 131 LIPDNSSLLKLCPSPEDEADLTNSGSSPSSEDDALPSGSPWRKKLRKKCKEKEKMEBFP 190
QY 187 DQDISPLPQPSRNKSRKHTEALQKLRVKNRQLDLRSCLSPKQHQSPALQSTDDDEVVLV 246
Db 191 DQDISPLPQPSRNKSRKHTEALQKLRVKNRQLDLRSCLSPKQHQSPALQSTDDDEVVLV 250
QY 247 EGPVLVPOSSRLFTLKIRCRADLRLPVRMSEPLQNVVDHMANHLGVSPNRILLFGESL 306
Db 251 EGPVLVPOSSRLFTLKIRCRADLRLPVRMSEPLQNVVDHMANHLGVSPNRILLFGESL 310
QY 307 SPTATPTTLKLGVAIDICVVLASSSEATETSQELRLRVQKQKHOMLEISLSPSPKLV 366
Db 311 SPTATPTTLKLGVAIDICVVLASSSEATETSQELRLRVQKQKHOMLEISLSPSPKLV 370
QY 367 LMSHYEAMGLSGHKLSPFFDGTGTLGSKELPADLGESGDLIEVWG 412
Db 371 LMSHYEAMGLSGHKLSPFFDGTGTLGSKELPADLGESGDLIEVWG 416

RESULT 3
Q9GLZ9 PRELIMINARY; PRT; 408 AA.
AC Q9GLZ9;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Hypothetical 44.6 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN PARIETAL LOBE;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBDJ databases.
DR EMBL; AB050511; BAB17279.1; -.
DR InterPro; IPR000626; Ubiquitin.
DR SMART; SM00213; UBQ; 2.
DR PROSITE; PS50053; UBIQUITIN_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 408 AA; 45580 MW; 5E0D6DD066AC2F24 CRC64;
```

Query Match 65.8%; Score 1382; DB 6; Length 408;
Best Local Similarity 70.1%; Pred. No. 9.6e-90;
Matches 293; Conservative 31; Mismatches 78; Indels 16; Gaps 5;

```
QY 1 MAEPLRGPRSGRGARRARGCPRARQSPARLIPDTVLVDVSDSEEVLEV-- 58
Db 1 MAEPLRGPRSGRGARRARGCPRARQSPARLIPDTVLVDVSDSEEVLEV-- 52
QY 59 ----ADPVEVPVAPAPAPQEQSDSDSGAAGGPAGAPRTLVRRRRRLLDPGEAPV 114
Db 53 ARCADEVEVAPSEPPGPVASRSDSDSGAAGGPAGAPPREPVRRRRLVLDPEAPLV 112
QY 115 PVYSGKVQSSNLIPDNSSLLKLCPEPEDEADLTNSGSSPSSEDDALPSGSPWRKKLRKK 174
Db 113 PVYSGKVQSSNLIPDNSSLLKLYPPGDEEVEADSSGLYHEGSPSP--GSPWTKLRTK 171
QY 175 CKKEKMEEFPPQDQISPLPQPSSRNKSRKHTEALQKLRVKNRQLDLRSCLSPKQHQSP 234
Db 172 -DKEEKTKTEILDNSPLSPSPSPRTKRGKTRALKKSEVKNRQLDLRSCLSPSPFPQ 230
```



```
RESULT 6
O35482 ID O35482 PRELIMINARY; PRT; 1072 AA.
AC O35482;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE High molecular-weight neurofilament.
GN NF-H.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91038277; PubMed=2230956;
RA Chin S.S., Liem R.K.;
RT "Transfected rat high-molecular-weight neurofilament (NF-H)
coassembles with vimentin in a predominantly nonphosphorylated form.";
RL J. Neurosci. 10:3714-3726 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Chin S.S.M., Liem R.K.H.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
DR EMBL; AF031879; AAB87068.1; -.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Coiled coil; Intermediate filament.
SQ SEQUENCE 1072 AA; 115349 MW; 89A146D457A4D78B CRC64;

Query Match 6.9%; Score 145.5; DB 11; Length 1072;
Best Local Similarity 22.1%; Pred. No. 0.066;
Matches 95; Conservative 63; Mismatches 166; Indels 105; Gaps 20;

QY 4 PLRGGRSGRGARRARGRCPRAROSPARLIPDTVLVLDVSDSEEV-----55
Db 613 PVEAKSPAEEKSPVSPG-EAKSPAEEKSPATVKSPVEAKSPAEEKSPVTVK 671
QY 56 --LEVADPVVPVARUPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRLLDPGEAPV 113
Db 672 SPAEAKSPVEV-----KSPASVSPSEAKSPAGAKSPAEEKSPVPAKSPAEEK- 725
QY 114 VPVYSGVQSSLLIPDNSSLLKLCSEPEDEADLTN-----SGSPSEDDA---LP 162
Db 726 -PAEAKSPAEEKSPAEEKS-----PAEAKSPAEEKSPVEKSPAEEKSPKESGLAE 779
QY 163 SGSPWRKLRKKEKEKKEEPPDDQDISLPQPS-SRNKSR-----KHTEALQKREVNK 217
Db 780 AKSPAEEKSPVKEIKPAEAKSPAEEKSPKESPEAKTLDVKSPEAKTPAKEEAK 839
QY 218 RLQDLRSCLSPKHQSPALQSTDDVVLVEGVPVLPQSSRLFTLKIRCRADLVRLP---VR 274
Db 840 RPADIR---SPEQVKSPEAKESK-----PEKEETREKVPAPKKEEVKSPVEEVK 886
QY 275 MSPLQNVDMANHLGVSNRILLFGESELSPTTPSTKLGVADIIDCVVLASSSEA 334
Db 887 AKEP-----PKV-----EEKTP-ATPKTE-----VKESKDEA 915
QY 335 TETSQELRLVQKQKHQMLIEISLSPDPLKVLMSHYEEAMGLSGHKLSPFFDGTLSGK 394
Db 916 PKEAQ---KPAEKEPELTE--KPKDSPGEAKKEEAKKAAPE-----E 956
QY 395 ELPADICLE 403
Db 957 ETPAKLGVK 965

RESULT 7
Q9XW25 ID Q9XW25 PRELIMINARY; PRT; 1634 AA.
AC Q9XW25;
```

```
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Y18D10A.1 protein.
GN Y18D10A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; AL034393; CAA22308.1; -.
DR InterPro; IPR000637; AT hook.
DR SMART; SM00384; AT hook; 1.
DR PROSITE; PS00354; HMGI_Y; UNKNOWN 1.
SQ SEQUENCE 1634 AA; 179059 MW; DC20372F3AAF48D0 CRC64;

Query Match 6.7%; Score 140; DB 5; Length 1634;
Best Local Similarity 22.0%; Pred. No. 0.28;
Matches 88; Conservative 54; Mismatches 152; Indels 106; Gaps 18;

QY 3 EPLRGGRSGRGARRARGRCPRAROSPARLIPDTVLVLDVSDSEEVLEVADP- 61
Db 413 EMMKGESVKEAARMRAEASAGRSQAPGAPAPA-----ASELQDPQDFGLSMSDFG 465
QY 62 -----VEVPV-----ARLPA-PAKPEQSDSDSEGAAGPAGAPRTLVRRR 102
Db 466 SDETEEHQKQEKHPIAMVTRRSARLSALPVPKKAASSK--MPPPPSPSPSTPGRRGR 523
QY 103 R-----RLLDGEPAPVPVYSGVQS-SLNLIPDNSSLLKLCSEPEDEA-DLTNSGSS 154
Db 524 RPRTLSTMSMEAPAAAVTPAPRGPRSRSAKAVSENTEPLSEAPSPVKRGRGRPRSRST 583
QY 155 PS-EDDALPSGFWKRLKKEKEKKEEPPDDQDISLPQPSRNKSRKHT--EALQK 211
Db 584 MSITDESEPTSTTAAKSKRAEEDDEEEDLKTNKP-EKPKKPSKTTEETVGDVLKK 642
QY 212 -LREYVK-----RLQDLRSCLSPKHQSPALQSTD-----DE 242
Db 643 RLRTAKTATVIHTPGPPLRTRKMRAPTAVTSSKKEPKNAGSADSSINEEHEDE 702
QY 243 VVLVEGPV--LPQSSRLFTLKIRCRADLVRLPVRMSEPLQNVDMANHLGVSPNRILL 300
Db 703 TWLIEEQTLDLFQQTSQEPRISSCGSEL-----LDEQFDASEEHSGLTVPS----- 747
QY 301 FGSELSPTATSTLKLGVADIIDCVLASSSEATETSQE 340
Db 748 --APELTKNPAPPV-----PEASEASAE 768

RESULT 8
Q90WA4 ID Q90WA4 PRELIMINARY; PRT; 2075 AA.
AC Q90WA4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Pericentriolar material 1 protein.
GN PCML.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
```

```

OX NCBI_TaxID=31033;
RN (1)
RN SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RA Bolland D.J., van Geel M., Carim Todd L., Beck A.F., Grewal P.K.,
RA van der Maarel S., Frants R.R., de Jong P.J., Hewitt J.E.;
RT "Comparative sequence analysis of an evolutionary chromosomal
RT breakpoint in human, mouse and Fugu rubripes.";
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF389403; AAK69357.1; -
SQ SEQUENCE 2075 AA; 229612 MW; 88A27A058F15045B CRC64;

Query Match 6.5%; Score 135.5; DB 13; Length 2075;
Best Local Similarity 22.4%; Pred. No. 0.78;
Matches 92; Conservative 48; Mismatches 125; Indels 145; Gaps 19;

QY 13 RCGRRARRRGARCGRCPRAROSPARL-----IPPTV-----LVLDVSDSDEVLVND 60
DB 652 RCGNSNNQGRINTCEINRRSAANLRSLNTSTTECQYNRDTRYNEVDEDEDVLDNEE 711
QY 61 PVEVVARLPAPAKPEQSDSDSEGAAEGPAGAPRTLVRRRRRLDPGEAPVPVYSGK 120
DB 712 GAR-----AAAAP-----DSEGS-----RSSLGNGEF-VHKYHQCK 744
QY 121 VOSSLNLIPDNSSLKLCPESEPEDEADLTNCGSSPSBEDAL-----PSGSPWRKK 170
DB 745 VKOKRLQLOELMAMVQ-----SDTD-----GTANEDDILHQPNNTRAGPKPSPREPG 794
QY 171 LAKCKEKEKMEFPDODISPLPQPSRNKGRKHTALQKREVNKRLQDLRSCISPQK 230
DB 795 LSKKA-REKLYEE-----KIRQOKELQKLDHDKRLTELOGKIRDLQ 836
QY 231 HQSPALQST-----DDEVLVVEGPVLPGSSRLFTLKIRGRADLVLRVPMSEPLQNVND 285
DB 837 WACPDQSSVCSTSGSGGLRKVPV-----TPVSIPEPVQAVAS 876
QY 286 MANHLGVSPNRILLFGESELSPTATPTLKGVADIIDCVLIASSSEATSTSOELR--- 342
DB 877 -----GPKRNVSL-----KKAAPPAVAPVAD-----NELMSEWRRIQ 910
QY 343 -LRVQCKEKNQMLETSLS-----PDSPLKVLMSHYEAMGLSGHKIS 383
DB 911 ILREBLRHRKRLLEYLIAHQRRSGPTDSP-----RHNKGLASPSHTVS 955

RESULT 9
Q9XVS4 PRELIMINARY; PRT; 971 AA.
AC Q9XVS4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE C25A1.10 protein.
GN C25A1.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodetidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RN SEQUENCE FROM N.A.
RA Mortimore B.J.;
RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
RN (2)
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; Z81038; CAB02755.1; -
DR InterPro; IPR002965; P_tich_extenm.
DR PRINTS; PRO1217; PRICHEXTENSN.
SQ SEQUENCE 971 AA; 101097 MW; 832936115B58709B CRC64;

```

```

Query Match 6.3%; Score 133; DB 5; Length 971;
Best Local Similarity 23.3%; Pred. No. 0.44;
Matches 60; Conservative 35; Mismatches 115; Indels 48; Gaps 8;

QY 29 PPAROSPARLIPDTLVLDVLS---DSDEVLVADPVEVPVAPRLAPAKPEDSDSG 85
DB 231 PAAKPTPAKPTPKPVAKAESSSDSDDEKKPVAKP--APAKPTPKPAKKADSSSDSD 288
QY 86 ----AAEGPAGAPRTLVRRRRRLD-----PGAPVPVYSGVQSSLNLI 128
DB 289 DEAPAKTTPAKAAPKPVAKKAESSSDSDDEKKPAKPTPAKATPKPVAKKAESSSDSD 348
QY 129 PDNSSLKLCSE--PEDEADLTNCGSSPSBEDALPGSPW-RKKLRKKCEKEMKEEF 185
DB 349 DEKKVPAPAKATPKPVAKKAESSSDSDDEKKPAKPTPAKATPKPVAKKAESSSDS 408
QY 186 PDQDISPLPQPSRNKS-----RKHTALOKLREVNKRLQDLNSC 225
DB 409 SDDEKKPVAKPTSAKATPKPAKKAADSSSDSDDEAPAKTTPAKAAPKASKAE---SS 465
QY 226 LSPKHOSPALQSTDEV 243
DB 466 SDSDEKPAKSTPAKI 483

RESULT 10
Q92541 PRELIMINARY; PRT; 664 AA.
AC Q92541;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE KIAA0252 protein (Fragment).
GN KIAA0252.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RN SEQUENCE FROM N.A.
RA TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Chara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and Brain.";
RL DNA Res. 3:321-329 (1996).
DR EMBL; D87440; BAA13382.1; -
DR InterPro; IPR004343; Plus-3.
DR Pfam; PF03126; Plus-3; 1.
FT NON-TER 1
SQ SEQUENCE 664 AA; 75809 MW; 2C6537A98BFBADEB CRC64;

Query Match 6.2%; Score 131; DB 4; Length 664;
Best Local Similarity 21.5%; Pred. No. 0.37;
Matches 101; Conservative 69; Mismatches 151; Indels 148; Gaps 19;

QY 37 RLIPDTLVLDVSDS--DEVLVLA-----DPVEVVARLPAPAKPEQSDSDSEGAA 87
DB 2 RVIDSDTEDSGSDNLQDELLSLAKRKRSDEKEPVSG-PAASPSSETSDSDDEWT- 59
QY 88 ECPGAPRTLVRRRRRLDPGEAPVNVVYSGKVOSSLNLIPDNSSLKLCSESEPEDEAD 147
DB 60 ---FGSNKNNKKKGRKRIEKG---TMKKQANKTASSGSDKD--SSASSAPEGEVSDS 112
QY 148 LTNCGSSPSBEDALPGSGPW----- 167
DB 113 DSNSSSSSDSDSDEDFHDGVEDLMGDEBDRAKLEQMTKEKREDELFRIRKREVL 172
QY 168 -----RKLR--KKCEKEMKEFPDQDISPL-----POPSSRNKR-----KH 205

```

Db 173 KRPFETKKLTKAKKKEKKKQEQEKKLTOIQESQVTSNHNKRRSKEDEKLDKX 232
QY 206 TEALQKLR-EVNKRLQDLRSLSPKQ-----HOSPALQSTD 240
Db 233 SQAMEELKAEREKKRNTAELLAKQPLKTSEVYSDDEEEEDDKSSEKSDRSRTSSD 292
QY 241 DEVVLVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSPNRIILL 300
Db 293 EEEKEEIP--PKSQ-----PVSPEEL-----NVRVLSRHKLRLW 326
QY 301 FGSELSPTATPOTKLIG-----VADIIDCVLASSSE--ATETSOELRLRVQ 347
Db 327 CHMPFFAKTVTGCFVIGIGNHNSKPVYVAEITGVETAKVYQLGCTRTNKLQRLRHGN 386
QY 348 KEKHQMLEISLSDPSLKVLMSHYEAMGLSGHKLJFFDGTGKLSQKEL 396
Db 387 DQVFRLEFVSNGEFTSEBFMK-WKEAMFSAGWQLPTL---DEINKKEL 431

RESULT 11
Q9H5F9 PRELIMINARY; PRT; 704 AA.
AC Q9H5F9;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-DEC-2001 (Tremblrel. 16, Last sequence update)
DE CDNA: FLJ23471 fis, clone HS11969.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK027124; BA015667.1; -
DR InterPro: IPR002965; P rich extensin.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 704 AA; 75768 MW; 858F94EEA2C1F8C6 CRC64;

Query Match 6.2%; Score 130; DB 4; Length 704;
Best Local Similarity 22.6%; Pred. No. 0.47;
Matches 100; Conservative 63; Mismatches 173; Indels 106; Gaps 20;

QY 10 PRSRGGRARRGARGRCPRARQSPARLIPDTVLVDVSDSEEVLEVADPV----- 62
Db 286 PKTEAPQASPLAKPLOSSSPRVGLPSRMEPPAPL---STSTSQASALPPAGRRLAE 341
QY 63 -----EVPVRLPAPAKPEQSDSDSEGAAGPAG-----AP 94
Db 342 SSGVGRVAGSRKPEAPMAKGSITLTQDMSTSLQEGDGPAGWRANLKPVDRRSPAE 401
QY 95 RTLVRRRRRLDP--GEAP--VVPVYSGKVQSSLLNI-PDNS-----SLLKLCPSSE 141
Db 402 RTLKPEPRALAPRAGEAPRKVSGFAGSVHITLTPVRPDRTPRASPSPSP 461
QY 142 PEDEADLTNSGSPSDEDDALPGSPGWRKRLKKEKEEK--MEEPDQDISP-----LP 194
Db 462 PR-RRRLAVPASLDVCDNWLRLPPEPPQEARVQSWKEEKKPHLQGRGPRPLSPANVPALP 520
QY 195 QPSRNRKSRKHTALOKLREVNRKLODLRSLSPKQHSQSPALQST-----DD--EVLVL 246
Db 521 GETVTSPVRLHPDLYSP-EIQQLQDIERRLDALRGVLEKRLRAAEGDAEDSLMW 579
QY 247 EGVLPQSSRLFTLKIRCRADLV-RLPVRMSEPLQNVVDHMANHLGVSPNRIILL---FG 302
Db 580 DWFNLTHEKQLL---LRQESLWYKSAQRLQEQDLIEGELRLMAKPEALKSLQERRR 636
QY 303 ESELSPTATPSTKLGVADIIDCVLASSSEATETSOELRLRVQKKEKHQMLEISLSPDS 362

Db 637 EQEL-----LEQVSTVND-----RSDIVSLDEDRLEQ--EEDQMLRMI----- 676
QY 363 PLKVLMSHYEAMGLSGHKLSP 384
Db 677 -----EKLGLQRKSKSF 688

RESULT 12
Q18866 PRELIMINARY; PRT; 504 AA.
AC Q18866;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Hypothetical 55.4 kDa protein.
GN C55C3.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Woessne J., Stellyes L.;
RT "The sequence of C. elegans cosmid C55C3.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53335; AL27231.1; -
KW Hypothetical protein.
SQ SEQUENCE 504 AA; 55405 MW; 9605686914F44659 CRC64;

Query Match 6.2%; Score 129.5; DB 5; Length 504;
Best Local Similarity 21.3%; Pred. No. 0.33;
Matches 91; Conservative 52; Mismatches 127; Indels 157; Gaps 21;

QY 11 RSRGGRARRGARGRCPRARQSPARLIPDTVLVDVSDSEEVLEVADPVVVARLP 70
Db 29 RRRGNGQIKKKPAK---PLKRITP-----NNSDEKLKKMKSP----- 63
QY 71 APAKPEQSDSDSEGAAGPAGAPRTLVRRRRRRLDP---GEAPVVPYSGKVQSSLLNL 127
Db 64 -PKPLQATDSD-----ILQPLPSNNAMTVP-----ESKESANR 96
QY 128 IPDNSSL-----LKLCPSEPE-----DEADLTNSGSPSDEDDALPGSPGWRKRLR 172
Db 97 SQSQSKINLNTDLDKDC-SEARAIPKCDKLTDFQTSTNSEQSPA--DAISAVAP-TKLGTG 152
QY 173 KKCEKEEKMEFPD-----QDI-----SPLQPSRRNK-----SRK 204
Db 153 SQSQSERQEIEKMPNSFRELODSKFKPAEAPIPKPEERKSGSTEQSTTPTKLQSGSPS 212
QY 205 HTEALQKLRVNRKLOD--LRSLSPKQHSQSPALQSTDD----- 241
Db 213 ERQEIQQMPNSFRELODSKFKPAPSPKSEKGAQSLSEPSVPGNNRRNSTSQVDITIS 272
QY 242 ----EVLVVEGVPVLPSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDH-----MANHLG 291
Db 273 PVPYTKLVGTQSPSERQEIEQMPNSFRELODSKFKPAQAPNPKPERVERGSAEQSMSEPLS 332

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 10:23:09 ; Search time 39 Seconds

(without alignments)
1407.673 Million cell updates/sec

Title: US-09-617-923-2

Perfect score: 2099
Sequence: 1 MAEPRLRGRRGRRGARR.....GKELPADLGESGDLIEWG 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID22/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:*
2: /SID22/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:*
3: /SID22/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:*
4: /SID22/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:*
5: /SID22/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:*
6: /SID22/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:*
7: /SID22/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:*
8: /SID22/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:*
9: /SID22/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:*
10: /SID22/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:*
11: /SID22/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:*
12: /SID22/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:*
13: /SID22/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:*
14: /SID22/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:*
15: /SID22/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:*
16: /SID22/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:*
17: /SID22/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:*
18: /SID22/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:*
19: /SID22/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:*
20: /SID22/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:*
21: /SID22/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:*
22: /SID22/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:*
23: /SID22/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2099	100.0	412	18 AAM34091	Mouse NF-AT intera
2	2099	100.0	412	20 AAY08331	Mouse NIP45 protei
3	1402.5	66.8	419	23 AAU10376	NF-AT interacting
4	1389.5	66.7	419	20 AAY08330	Human NIP45 protei
5	1399.5	66.7	419	22 AAM93780	Human polypeptide,
6	1079	51.4	303	23 AAU10374	NF-AT interacting
7	987	47.0	334	23 AAU10372	NF-AT interacting
8	785.5	37.4	286	23 AAU10373	NF-AT interacting
9	623	29.7	138	22 AAM94004	Human stomach can
10	623	29.7	138	22 AAB94289	Human protein seq.

11	255	12.1	55	23 AAU10375	NF-AT interacting
12	235	11.2	87	23 AAO14942	Human NF-AT intera
13	154.5	7.4	1219	22 ABG12658	Novel human diago
14	154.5	7.4	1279	22 AAM78969	Human protein SEQ
15	154.5	7.4	1301	22 AAM79953	Human protein SEQ
16	148	7.1	710	22 ABG20363	Novel human diago
17	143.5	6.8	2013	22 ABG12657	Novel human diago
18	137	6.5	990	23 AAE22762	Rat PGC-3 protein.
19	134	6.4	346	23 ABB06133	Human NS protein s
20	132	6.3	586	18 AAM10423	Cytomegalovirus UL
21	128.5	6.1	755	22 ABB59259	Drosophila melanog
22	128	6.1	281	22 ABG05583	Novel human diago
23	128	6.1	281	22 ABG18846	Novel human diago
24	127.5	6.1	1167	22 ABB57757	Drosophila melanog
25	127.5	6.1	1408	22 ABB58704	Drosophila melanog
26	127	6.1	1368	22 AAB60262	Drosophila melanog
27	126.5	6.0	676	22 AAM47218	Human NOV9 protein
28	126.5	6.0	676	22 AAE09441	Human SB-HDGF prot
29	126.5	6.0	676	22 AAG64373	Human hepatome cel
30	126.5	6.0	676	23 ABB05039	Human NOV4a protei
31	126.5	6.0	717	22 AAE09442	Human SB-HDGF prot
32	125.5	6.0	1014	23 AAU96745	PCG2-PPAR gamma co
33	124.5	5.9	518	21 AAB53332	Human colon cancer
34	124.5	5.9	555	22 AAG75075	Human colon cancer
35	124.5	5.9	671	21 AAY9426	Human PDI604 (UNQ
36	124.5	5.9	671	22 AAU29196	Human PRO polypept
37	124.5	5.9	671	22 AAB66175	Protein of the inv
38	124.5	5.9	671	23 ABB05042	Human NOV4d protei
39	124.5	5.9	1783	23 ABB97448	Novel human protei
40	124	5.9	1191	22 AAM80219	Human protein SEQ
41	123.5	5.9	671	23 AAY97038	Human LP protein L
42	123	5.9	950	20 AAY33298	Human membrane spa
43	122.5	5.8	1152	22 ABB57951	Drosophila melanog
44	122.5	5.8	1152	22 ABB66110	Drosophila melanog
45	122.5	5.8	1152	22 ABB67224	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAW34091	AAW34091 standard; Protein; 412 AA.
XX	AAW34091;
XX	18-MAY-1998 (first entry)
XX	Mouse NF-AT interacting protein 45.
DE	
XX	NF-AT Interacting Protein 45; NIP45; yeast two-hybrid assay; mouse;
KW	Rel homology domain; RHD; T cell; transcription factor; cancer;
KW	interleukin-4; IL-4; development; Th1; Th2; cytokine; allergy;
KW	autoimmune disease; transplantation.
XX	
OS	Mus sp.
XX	
FH	
FT	Key
FT	Region
XX	
PN	MO9739721-A2.
XX	
PD	30-OCT-1997.
XX	
PF	23-APR-1997; 97WO-US06708.
XX	
PR	25-NOV-1996; 96US-0755592.
PR	23-APR-1996; 96US-0636602.
PR	25-NOV-1996; 96US-0755584.
XX	
PA	(HARD) HARVARD COLLEGE.
XX	

PI Glincher LH, Ho I, Hodge MR;
XX WPI; 1997-535556/49.
DR N-PSDB; AAT93045.
XX Production of cytokine(s) associated with Th2-type helper T cells -
PT particularly for controlling development of Th1 and Th2 cells for
PT treatment of allergy, autoimmune disease etc.
XX Claim 31; Fig 11; 151pp; English.
XX This is the amino acid sequence of the mouse NF-AT Interacting Protein 45
CC (NIP45). The gene sequence was isolated by using a yeast two-hybrid
CC detection system for proteins that interact with the NF-AT Rel homology
CC domain (RHD). The assay used, as a "bait", a 900 bp fragment of the
CC murine NF-ATp encoding the region spanning amino acids 228-250. NF-AT is
CC a multisubunit transcription complex containing a cyclosporin A sensitive
CC cytoplasmic phosphoprotein and an inducible component of the AP-1 family
CC of transcription factors. The screen was carried out on a cDNA library
CC prepared from the murine T cell line D10. One class of proteins,
CC designated NIP45, bound the NF-AT-RHD region with high affinity. NIP45
CC can be used in a claimed method to inhibit or stimulate production of
CC NF-AT family protein, particularly interleukin-4 (IL-4) and the treated
CC cells may be administered to control development of Th1 or Th2 cells by
CC modulating the production of a T-helper type 2 associated cytokine.
CC Especially the method is useful in the inhibition of Th2 in allergy,
CC cancer or infections, and promotion of Th2 in autoimmune disease and
CC transplantation.
XX Sequence 412 AA;
SQ
Query Match 100.0%; Score 2099; DB 18; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.5e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDSEVLEVAD 60
DB 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDSEVLEVAD 60
QY 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120
DB 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120
QY 121 VOSSLNLIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180
DB 121 VOSSLNLIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180
QY 181 KMEFPDQDISPLPQSSRNKSRKHTALQKLRVKNRLQDLRSLCSPKHQSALQSTD 240
DB 181 KMEFPDQDISPLPQSSRNKSRKHTALQKLRVKNRLQDLRSLCSPKHQSALQSTD 240
QY 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
DB 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
QY 301 FGSELSPTATPSTTLKLGVAIDIICVVLASSSEATSTQELRLRVQKQKHQMLEISLSP 360
DB 301 FGSELSPTATPSTTLKLGVAIDIICVVLASSSEATSTQELRLRVQKQKHQMLEISLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412
DB 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412
RESULT 2
AAY08331
ID AAY08331 standard; Protein; 412 AA.
XX
AC AAY08331;
XX
DT 16-JUL-1999 (first entry)
XX
DE Mouse NIP45 protein.

XX NIP45; mouse; transcription transactivator; IL-4; interleukin-4;
KW NF-AT interacting protein-45; modulator; treatment; inflammation;
KW autoimmune disease; HTP screening; drug testing; allergic; therapy;
KW T-cell dependent response; cytokine; diagnostic; immunosuppressant.
XX Mus musculus.
XX WO9921993-A1.
XX 06-MAY-1999.
XX 21-OCT-1998; 98WO-GB03141.
XX 24-OCT-1997; 97GB-0022388.
XX (ZENE) ZENECA LTD.
XX Liu D, Zhao J, Zhou H;
XX WPI; 1999-312964/26.
XX Polynucleotides encoding human NF-AT interacting protein (NIP45)
XX Disclosure; Fig 4; 84pp; English.
XX This invention describes a novel human NF-AT interacting protein-45
CC NIP-45. Human NIP45 is a transcriptional trans-activation factor of the
CC interleukin 4 (IL-4) gene. Modulators of human NIP45 and IL-4 can be used
CC to treat conditions mediated by NIP45 or IL-4, e.g. inflammation or
CC autoimmune disease. NIP45 is a good candidate target for HTP screening
CC and/or testing system for drugs that will alleviate T-cell dependent
CC autoimmune and allergic responses, and for cytokine-based therapies of
CC chronic disease. Expression of NIP45 can be inhibited, and IL-4
CC expression modulated, by administering an antisense molecule. Antibodies
CC against NIP45 and primers can be used in diagnostic assays. Discovery of
CC a human NIP45 is advantageous in that it provides the ability to control
CC IL-4 transcription, which is of importance for anti-inflammation and
CC immunosuppressant drug development.
XX Sequence 412 AA;
SQ
Query Match 100.0%; Score 2099; DB 20; Length 412;
Best Local Similarity 100.0%; Pred. No. 1.5e-174;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDSEVLEVAD 60
DB 1 MAEPLRGPRSGRGARRARGRCPRARQSPARLIPDTVLVDLVSDSDSEVLEVAD 60
QY 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120
DB 61 PVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLLDPGEAPVPVYSGK 120
QY 121 VOSSLNLIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180
DB 121 VOSSLNLIPDNSSLLKLCFSEPEDEADLTNSGSSPSEDDALPSGSPWRKKLRKCKEKEK 180
QY 181 KMEFPDQDISPLPQSSRNKSRKHTALQKLRVKNRLQDLRSLCSPKHQSALQSTD 240
DB 181 KMEFPDQDISPLPQSSRNKSRKHTALQKLRVKNRLQDLRSLCSPKHQSALQSTD 240
QY 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
DB 241 DEVVLVEGVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHLGVSFNRILL 300
QY 301 FGSELSPTATPSTTLKLGVAIDIICVVLASSSEATSTQELRLRVQKQKHQMLEISLSP 360
DB 301 FGSELSPTATPSTTLKLGVAIDIICVVLASSSEATSTQELRLRVQKQKHQMLEISLSP 360
QY 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412
DB 361 DSPLKVLMSHYEAMGLSGHKLSPFFDGTGKLSGKELPADLGLESGLIEVWG 412

RESULT 3
 ID AAU10376 standard; Protein; 419 AA.
 AC AAU10376;
 DT 14-FEB-2002 (first entry)
 DE NF-AT interacting protein 45, NIP45.
 XX NF-AT interacting protein 45 variant; NIP45V, immunosuppressive; human;
 KW antiallergic; antiaesthetic; antigout; dermatological; antiinflammatory;
 KW antipruritic; antibacterial; vasotropic; autoimmune; allergy; asthma;
 KW nuclear factor of activated T cells; chronic inflammatory disease;
 KW allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;
 KW vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;
 KW diarrheic colitis; polymyositis; lymphofollicular thymitis; gout;
 KW dermatomyositis; Sjogren's syndrome.
 OS Homo sapiens.
 PN WO200181574-A2.
 PD 01-NOV-2001.
 PE 25-APR-2001; 2001WO-EP04635.
 PR 25-APR-2000; 2000US-199356P.
 PA (FARB) BAYER AG.
 PI Encinae J;
 DR WPI; 2002-05355/07.
 DR N-PSDB; AAS17539, AAS17540.
 XX Novel nuclear factor of activated T cells interacting protein NIP45
 PT variants which can be regulated for treating conditions involving
 PT inflammatory processes such as allergy, asthma and autoimmune diseases
 PT -
 PS Disclosure; Fig 12; 107bp; English.
 XX
 XX The invention relates to an isolated nuclear factor of activated T cells
 CC (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and
 CC polynucleotide (II) encoding (I). (I) and (II) are useful for screening
 CC for agents which decrease the activity of NIP45V polypeptide. (I) is also
 CC useful for screening for agents which regulate the activity, preferably
 CC cyclic AMP formation, mobilisation of intracellular calcium or
 CC phosphoinositide metabolism of human NIP45V. (II) is useful for detecting
 CC a polynucleotide encoding a NIP45V polypeptide in a biological sample, by
 CC hybridising the polynucleotide to a nucleic acid material of a biological
 CC sample. Antibody which specifically binds to (I) or (II) is useful for
 CC reducing the activity of NIP45V in a cell, and for detecting (I) in
 CC a sample. (I) is useful for modulating the activity of NIP45V in a
 CC disease, such as autoimmune, allergic, infectious or chronic inflammatory
 CC disease or asthma. NIP45V is useful for controlling diseases including
 CC allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal
 CC catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,
 CC diarrheic colitis, systemic inflammatory response syndrome, polymyositis,
 CC lymphofollicular thymitis, sepsis, dermatomyositis, mixed connective
 CC tissue disease (MCTD), Sjogren's syndrome and gout. (II) is also useful
 CC in diagnostic assays for detecting diseases and abnormalities or
 CC susceptibility to diseases related to the presence of mutations in
 CC nucleic acid sequence which encode (I). (I) or its fragment is useful
 CC in raising specific antibodies. The present sequence represents the
 CC amino acid sequence of NIP45 as described in the invention.
 XX
 XX Sequence 419 AA;

	Best Local Similarity	70.1%	Pred. No. 9.5e-114:	
	Matches	295;	Conservative	30; Mismatches 85; Indels 11; Gaps 4
QY	1	MAEPIRGGRPRG---	GRGARARGARCRCPRARQSPARLIPDTVLVDVSDSEEVLE	57
Db	1	MAEPVGRKRWGGSGGARGRGGRGGRAPRAGRSPRGTLVDVVSVDLVSDSEELIE	60	
QY	58	V-----ADPVEPVARLLPAPAKPQDSDSDSEGAAGPAGPRTLVRRRRLLDPGEA	111	
Db	61	VATARGADEVEVEPEPPPGPVASRDNSNSDEGDRRACGPPRRPVRRRLRLVLDPGEA	120	
QY	112	PVVPVYSGVQSSLLIPDSSLLRLCSEPEDEADLTNMGSSPSEDDALPFGSPWRKYL	171	
Db	121	PLVPVYSGKVASLLLIPDSSLLRLYPGDEEBEALADSSGLYHGGSPSP-GSPMKTKYL	179	
QY	172	RKKCKEKKKEEFPDODISPLPQPSNRKSRKRTKATLQKLEVKRRLQDLRSLCSPKOH	231	
Db	180	RKKCKEKKKEFLLDLDNSPLSPSPSPKSTHTRALKUSEVKRRLQDLRSLCSLSPKPP	238	
QY	232	QSPALQSTDDVEVLYEGPYLPOSSRLFTLTKRCRADLVRLPRMSEPLQNVVDHMANHIG	291	
Db	239	QCOEOQOGEDEVLYEGPRTLPTPTPLPLKTRCRADLVRLPRMSEPLQSVVDHMAHATIG	298	
QY	292	VSPNRLLFGSESELPTATPESTLRLGVAADIDCVLASSPSEATSOELRLRYOGKEKH	351	
Db	299	VSPSILLFGSTELSPATPTPLRLGVAADIDCVLVSSPATETSOQLQRLRYOGKEKH	358	
QY	352	QMLLEISLPDSELUKLYLMHYEAMQLSGHKLSFFDGTLSGKEPLPADIGLESGDLIEWW	411	
Db	359	QTLVEVSLSDSLDKLTMHGYEAMQLSGRKLSPFDGTLSGRELPLADIGMSSGDLIEWW	418	
QY	412	G	412	
Db	419	G	419	

CC	This invention describes a novel human NF-AT interacting protein-45
PS	Claim 1; Fig 3; 84pp; English.
PT	Polynucleotides encoding human NF-AT interacting protein (NIP45)
XX	
DR	N-PSDB; AAX56955.
XX	
DR	WI; 1999-312964/26.
XX	
PI	Liu D, Zhao J, Zhou H;
PA	(ZENECA LTD.
XX	
PR	24-OCT-1997; 97GB-0023288.
XX	
PF	21-OCT-1998; 98WO-GB03141.
XX	
PD	06-MAY-1999.
XX	
PN	WO9921993-A1.
XX	
OS	Homo sapiens.
XX	
KW	T-cell dependent response; cytokine; diagnostic; immunosuppressant.
KM	autoimmune disease; HTP screening; drug testing; allergic therapy;
KW	NF-AT interacting protein-45; modulator; treatment; inflammation;
KM	IL-4; interleukin-4;
KW	Human NIP45 protein.
DE	
DT	16-JUL-1999 (first entry)
XX	
AC	AAV08330;
XX	
ID	AAV08330 standard; Protein; 419 AA.
XX	
AY08330	
RESULT 4	


```

XX 14-FEB-2002 (first entry)
XX NF-AT interacting protein 45 variant 3, NIP45V1.
XX
XX NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;
XX antiallergic; antiasthmatic; antigout; dermatological; antiinflammatory;
XX antipsoarctic; antibacterial; vasotrophic; autoimmune; allergy; asthma;
XX nuclear factor of activated T cells; chronic inflammatory disease;
XX allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;
XX vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;
XX diarrhetic colitis; polymyositis; lymphofollicular thyroiditis; gout;
XX dermatomyositis; Sjogren's syndrome.
XX Homo sapiens.
XX WO200181574-A2.
XX
XX 01-NOV-2001.
XX
XX 25-APR-2001; 2001WO-EP04635.
XX
XX 25-APR-2000; 2000US-199356P.
XX
XX (FARB ) BAYER AG.
XX
XX Encinas J;
XX
XX WPI; 2002-055355/07.
XX
XX N-PSDB; AAS17537.
XX
XX Novel nuclear factor of activated T cells interacting protein NIP45
XX variants which can be regulated for treating conditions involving
XX inflammatory processes such as allergy, asthma and autoimmune diseases
XX
XX
XX Claim 1; Fig 10; 107pp; English.
XX
XX The invention relates to an isolated nuclear factor of activated T cells
XX (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and
XX polynucleotide (II) encoding (I). (I) and (II) are useful for screening
XX for agents which decrease the activity of NIP45V polypeptide. (I) is also
XX useful for screening for agents which regulate the activity, preferably
XX cyclic AMP formation, mobilisation of intracellular calcium or
XX phosphoinositide metabolism of human NIP45V. (II) is useful for detecting
XX a polynucleotide encoding a NIP45V polypeptide in a biological sample, by
XX hybridising the polynucleotide to a nucleic acid material of a biological
XX sample. Antibody which specifically binds to (I) or (II) is useful for
XX reducing the activity of NIP45V in a cell, and for detecting (I) in
XX a sample. (I) is useful for modulating the activity of NIP45V in a
XX disease, such as autoimmune, allergic, infectious or chronic inflammatory
XX disease or asthma. NIP45V is useful for controlling diseases including
XX allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal
XX catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,
XX diarrhetic colitis, systemic inflammatory response syndrome, polymyositis,
XX lymphofollicular thyroiditis, sepsis, dermatomyositis, mixed connective
XX tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful
XX in diagnostic assays for detecting diseases and abnormalities or
XX susceptibility to diseases related to the presence of mutations in
XX nucleic acid sequence which encode (I). (I) or its fragment is useful
XX in raising specific antibodies. The present sequence represents the
XX amino acid sequence of NIP45V as described in the invention.
XX
XX Sequence 303 AA;
XX
XX Query March 51.4%; Score 1079; DB 23; Length 303;
XX Best Local Similarity 75.4%; Pred. No. 1e-85;
XX Matches 221; Conservative 21; Mismatches 49; Indels 2; Gaps 2;
XX
XX 120 KVGSSINLIPNSSLKLCPESEADLTNSGSSPEDDLPSGPKRLKRRKCEKEE 179
XX :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
XX 13 EVKSLRIIPDLSTLKKLYPGEDEEALADSSGLYHGSGSP-GSPKTKLRITK-DKEE 70

```

```

Qy 180 KMEEFPPQDLSPLPQESSRNKSRKHTALOKLREVNKRLQDLNSCLSPKQHQSPALQST 239
Db 71 KKTEFFLDLNSPLSPSPSPRTKSRHTRLAKLSEVNRRLDLSRCLSPKQEQOQO 130
Qy 240 DDEVVLVEGCVLPQSSRFLTKIRCRADLVRLPVRMSEPLQNVDMHNGVSPNRLTL 299
Db 131 EDEVVLVEGCVLPETPRFPPLKIRCRADLVRLPLRMSEPLQSVDDHMTGIVGSPSRLTL 190
Qy 300 LFGSELSPTATPSTLKLGVADIIDCVLASSSEATETSOELRLRVQGEKRMLEISLS 359
Db 191 LFGTELSPTATPRTLKGVADIIDCVLTSSPEATETSOQLQRLVQGEKRGHTLEVLSS 250
Qy 360 PDSPLKVLMSHYEAMGLSGHKLSPFDGTLKSGKELPADLGLBSGDLIEVWG 412
Db 251 RDSPLKTLMSHYEAMGLSGRKLSPFFDGTLSGRELPA DLGMESGDLIEVWG 303
RESULT 7
AAU10372
ID AAU10372 standard; Protein; 334 AA.
XX
XX AAU10372;
XX
XX 14-FEB-2002 (first entry)
XX
XX NF-AT interacting protein 45 variant 1, NIP45V1.
XX
XX NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;
XX antiallergic; antiasthmatic; antigout; dermatological; antiinflammatory;
XX antipsoarctic; antibacterial; vasotrophic; autoimmune; allergy; asthma;
XX nuclear factor of activated T cells; chronic inflammatory disease;
XX allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;
XX vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;
XX diarrhetic colitis; polymyositis; lymphofollicular thyroiditis; gout;
XX dermatomyositis; Sjogren's syndrome.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX MISC-difference 334
XX /note= "Encoded by gc"
XX
XX WO200181574-A2.
XX
XX 01-NOV-2001.
XX
XX 25-APR-2001; 2001WO-EP04635.
XX
XX 25-APR-2000; 2000US-199356P.
XX
XX (FARB ) BAYER AG.
XX
XX Encinas J;
XX
XX WPI; 2002-055355/07.
XX
XX N-PSDB; AAS17534, AAS17535.
XX
XX Novel nuclear factor of activated T cells interacting protein NIP45
XX variants which can be regulated for treating conditions involving
XX inflammatory processes such as allergy, asthma and autoimmune diseases
XX
XX Claim 1; Fig 8; 107pp; English.
XX
XX The invention relates to an isolated nuclear factor of activated T cells
XX (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and
XX polynucleotide (II) encoding (I). (I) and (II) are useful for screening
XX for agents which decrease the activity of NIP45V polypeptide. (I) is also
XX useful for screening for agents which regulate the activity, preferably
XX cyclic AMP formation, mobilisation of intracellular calcium or
XX phosphoinositide metabolism of human NIP45V. (II) is useful for detecting
XX a polynucleotide encoding a NIP45V polypeptide in a biological sample, by
XX hybridising the polynucleotide to a nucleic acid material of a biological

```

CC sample. Antibody which specifically binds to (I) or (II) is useful for
CC reducing the activity of NIP45V in a cell, and for detecting (I) in
CC a sample. (I) is useful for modulating the activity of NIP45V in a
CC disease, such as autoimmune, allergic, infectious or chronic inflammatory
CC disease or asthma. NIP45V is useful for controlling diseases including
CC allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal
CC catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,
CC diabetotic colitis, systemic inflammatory response syndrome, polymyositis,
CC lymphofollicular thymitis, sepsis, dermatomyositis, mixed connective
CC tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful
CC in diagnostic assays for detecting diseases and abnormalities or
CC susceptibility to diseases related to the presence of mutations in
CC nucleic acid sequence which encode (I). (I) or its fragment is useful
CC in raising specific antibodies. The present sequence represents the
CC amino acid sequence of NIP45V1 as described in the invention.
XX
SQ Sequence 334 AA;

Query Match 47.0%; Score 987; DB 23; Length 334;
Best Local Similarity 52.7%; Pred. No. 1.3e-77;
Matches 222; Conservative 24; Mismatches 79; Indels 96; Gaps 5;

QY 1 MAEPLRGGRPRSG---GRGARRARGRCPRARQSPARLIPDTVLVLDVSDSEVLE 57
DB 1 MAEPVGRGRWGGSGAGRGGRGGGRAPRAQRSPSRGTLVSVDLVTDSEILE 60

QY 58 V-----ADPEVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEA 111
DB 61 VATARGAADEVEPEPPPGPVASRDNSDSEGEDRRRPPAGPREPVRRRRLVLDPGEA 120

QY 112 PVVTVSGKVQSSNLIPDNLKCPSEPEDEADLTNSGSSPSDDALPSGSPWRKKL 171
DB 121 PLVTVSGKVKSRLIPDLSLLKLYPPGDEEAEALDSSGLYHEGSPSP-GSPWKTL 179

QY 172 RKKEEKKMEFFPDQDISPLPQPSNRNKRKHTALQKLRVKNRQLDLRSLSPKQH 231
DB 180 RTK-DKEEKKTEFLDNLSPSPSPRSTKSRTHTRALKKLVSEVKNRQLDLRSLSPKPP 238

QY 232 QSPALQSTDDVVLVEGVPVLPQSSRLTKIRCRADLVRLPVMSEPLQNVVDHMANHLG 291
DB 239 QGQEQGQDEVLVEGTLPTPRFLPKIRCRADLVRLPLRM----- 282

QY 292 VSPNRILLFGSELSPTATPSTLKLGVADIIDCVVLASSSEATSTQELRLRVOGKEKH 351
DB 283 ----- 282

QY 352 QMLEISLSPSPKVLMSHYEAMGLSGHKLSFFFDGTLGSKELPADLGESGLIEVW 411
DB 283 -----DSPLKTLMSHYEAMGLSGRKLSFFFDGTLGSKELPADLGESGLIEVW 333

QY 412 G 412
DB 334 G 334

RESULT 8
AAU10373
ID AAU10373 standard; Protein; 286 AA.
XX
AC AAU10373;
XX
DT 14-FEB-2002 (first entry)
XX
DE NF-AT interacting protein 45 variant 2, NIP45V2.
XX
KW NF-AT interacting protein 45 variant; NIP45V; immunosuppressive; human;
KW allergic; antiasthmatic; antigout; dermatological; antiinflammatory;
KW antipsoriatic; antibacterial; vasotropic; autoimmune; allergy; asthma;
KW nuclear factor of activated T cells; chronic inflammatory disease;
KW allergic rhinitis; atopic dermatitis; hives; conjunctivitis; psoriasis;
KW vernal catarrh; systemic lupus erythematosus; myasthenia gravis; sepsis;
KW diabetotic colitis; polymyositis; lymphofollicular thymitis; gout;
KW dermatomyositis; Sjogren's syndrome.

XX Homo sapiens.
XX OS
XX PN WO200181574-A2.
XX PD 01-NOV-2001.
XX PF 25-APR-2001; 2001WO-EP04635.
XX PR 25-APR-2000; 2000US-199356P.
XX PA (PARB) BAYER AG.
XX PI Encinas J;
XX DR WPI; 2002-055355/07.
XX DR N-PSDB; AAS17536.
XX
PT Novel nuclear factor of activated T cells interacting protein NIP45
PT variants which can be regulated for treating conditions involving
PT inflammatory processes such as allergy, asthma and autoimmune diseases
PT
XX Claim 1; Fig 9; 107pp; English.
XX
CC The invention relates to an isolated nuclear factor of activated T cells
CC (NF-AT) interacting protein NIP45 variant (NIP45V) polypeptide (I) and
CC polynucleotide (II) encoding (I). (I) and (II) are useful for screening
CC for agents which decrease the activity of NIP45V polypeptide. (I) is also
CC useful for screening for agents which regulate the activity, preferably
CC cyclic AMP formation, mobilisation of intracellular calcium or
CC phosphoinositide metabolism of human NIP45V. (II) is useful for detecting
CC a polynucleotide encoding a NIP45V polypeptide in a biological sample, by
CC hybridising the polynucleotide to a nucleic acid material of a biological
CC sample. Antibody which specifically binds to (I) or (II) is useful for
CC reducing the activity of NIP45V in a cell, and for detecting (I) in
CC a sample. (I) is useful for modulating the activity of NIP45V in a
CC disease, such as autoimmune, allergic, infectious or chronic inflammatory
CC disease or asthma. NIP45V is useful for controlling diseases including
CC allergic rhinitis, atopic dermatitis, hives, conjunctivitis, vernal
CC catarrh, systemic lupus erythematosus, myasthenia gravis, psoriasis,
CC diabetotic colitis, systemic inflammatory response syndrome, polymyositis,
CC lymphofollicular thymitis, sepsis, dermatomyositis, mixed connective
CC tissue disease (MCTD), Sjogren's syndrome and gout. (I) is also useful
CC in diagnostic assays for detecting diseases and abnormalities or
CC susceptibility to diseases related to the presence of mutations in
CC nucleic acid sequence which encode (I). (I) or its fragment is useful
CC in raising specific antibodies. The present sequence represents the
CC amino acid sequence of NIP45V2 as described in the invention.
XX
SQ Sequence 286 AA;

Query Match 37.4%; Score 785.5; DB 23; Length 286;
Best Local Similarity 61.1%; Pred. No. 3.9e-77;
Matches 174; Conservative 23; Mismatches 60; Indels 11; Gaps 4;

QY 1 MAEPLRGGRPRSG---GRGARRARGRCPRARQSPARLIPDTVLVLDVSDSEVLE 57
DB 1 MAEPVGRGRWGGSGAGRGGRGGGRAPRAQRSPSRGTLVSVDLVTDSEILE 60

QY 58 V-----ADPEVEVPVARLPAPAKPEQSDSDSEGAAGPAGAPRTLVRRRRRLDPGEA 111
DB 61 VATARGAADEVEPEPPPGPVASRDNSDSEGEDRRRPPAGPREPVRRRRLVLDPGEA 120

QY 112 PVVTVSGKVQSSNLIPDNLKCPSEPEDEADLTNSGSSPSDDALPSGSPWRKKL 171
DB 121 PLVTVSGKVKSRLIPDLSLLKLYPPGDEEAEALDSSGLYHEGSPSP-GSPWKTL 179

QY 172 RKKEEKKMEFFPDQDISPLPQPSNRNKRKHTALQKLRVKNRQLDLRSLSPKQH 231
DB 180 RTK-DKEEKKTEFLDNLSPSPSPRSTKSRTHTRALKKLVSEVKNRQLDLRSLSPKPP 238

QY 232 QSPALQSTDDVVLVEGVPVLPQSSRLTKIRCRADLVRLPVRMS 276

Db 239 QGQGGQGGDEVLVEGPTLPETPRLPPLKIRCADLVLLPRLMT 283

RESULT 9
ID AAM94004
XN AAM94004 standard; Protein: 138 AA.
XX
AC AAM94004;
XX
DT 13-NOV-2001 (first entry)
XX
DE Human stomach cancer expressed polypeptide SEQ ID NO 78.
XX
KW Human; stomach cancer; marker; screening; micro-metastasis;
KM peritoneal dissemination.
XX
OS Homo sapiens.
XX
PN WO200109317-A1.
XX
PD 08-FEB-2001.
XX
PF 28-JUL-2000; 2000WO-JP05063.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 18-OCT-1999; 99US-0159590.
PR 11-JAN-2000; 2000JP-0118776.
PR 17-FEB-2000; 2000US-0183322.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;
PI Kodama T, Midorikawa Y;
XX
DR WPI: 2001-570287/64.
DR N-PSDB: AAI93881.
XX
PT New Stomach cancer-associated genes, useful as markers in blood tests
PT for screening for the early stages of the disease -
XX
XX Claim 1; Page 150-151; 242pp; Japanese.
XX
PS The invention relates to stomach cancer-expressed genes
CC (AAI93842-AAI93917) and the encoded proteins (AAM93967-AAM94039). The
CC genes can be used as markers in blood tests for screening for the early
CC stages of the disease. The proteins and peptides can be used as targets
CC for screening for compounds to treat the disease. They can also be used
CC for predicting micro-metastases. The gene can predict peritoneal
CC dissemination.
XX
SQ Sequence 138 AA;
Query Match 29.7%; Score 623; DB 22; Length 138;
Best Local Similarity 88.4%; Pred. No. 2.1e-46;
Matches 122; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 275 MSEPLQNVVDHMANHLGVSPNRIILFGESLSPTATPSTIKLGVADIIDCVVLASSSEA 334
Db 1 MSEPLQSVVDHMAVTHLGVSPSRILLFGETELSPATPRTIKLGVADIIDCVVLASSSEA 60

QY 335 TETSOELRLRVOGKREKQMLISLSPDSPKLVLNMSHYEAMGLSGKLSFFPDGFKLSGK 394
Db 61 TETSOQOLRLRVOGKREKQMLISLSPDSPKLVLNMSHYEAMGLSGKLSFFPDGFKLSGR 120

QY 395 ELPADLGLSGDLIEVWG 412
Db 121 ELPADLGLSGDLIEVWG 138

RESULT 10
ID AAB94289
XN AAB94289 standard; Protein: 138 AA.
XX
AC AAB94289;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14733.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 8; SEQ ID 14733; 2537pp + CD ROM; English.
XX
PS The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAB92446 to AAB13628 and
CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAB13629 to AAB13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 138 AA;
Query Match 29.7%; Score 623; DB 22; Length 138;
Best Local Similarity 88.4%; Pred. No. 2.1e-46;
Matches 122; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 275 MSEPLQNVVDHMANHLGVSPNRIILFGESLSPTATPSTIKLGVADIIDCVVLASSSEA 334
Db 1 MSEPLQSVVDHMAVTHLGVSPSRILLFGETELSPATPRTIKLGVADIIDCVVLASSSEA 60

PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK52102.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 3969-3972; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1279 AA;
SQ
Query Match 7.4%; Score 154.5; DB 22; Length 1279;
Best Local Similarity 23.5%; Pred. No. 0.00042; Indels 129; Gaps 21;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;
QY 6 RRGGRSRGGRGARRGARGRCPRARQSPARLIPDTVLVDSDEEVLEADVPEVP 65
DB 515 RRGGRSQGFPSSQFQASG-----PNRHQAPSMLSFGPA---LSSDSKKEGEDEGTEELP 566
QY 66 -----VARLPAPAKPEQSDSDSEGAEE-----GPAGAPRTLVRRRRRLL 106
DB 567 ALPVLAKSTKKALASVSPALPRSLSHWMSRAQESVGFLLDPAPAAPNPGP---RRGRWV 623
QY 107 DRGEAPVVPYSGVQSLLNLPDNLSSLLKCPB--EP-EDEADLTNSGSSPSEDDALPS 163
DB 624 QPG-----VELSVRSMIDLRLQLETAPLSQDPSQDSLAIIPSGPRKHGQEALET 672
QY 164 G-----SPWRKKLRKK-----EKEE-----KMEEFPPQDITSPLPQPS---SRNK 201
DB 673 SLTSONEKPRPQASQPCSPHILRLLSQEGVFAQDLEPAPIEDGIVPEPDNDTMDT 732
QY 202 SRKHTEALQK-----LREVNKLQD-----LRCLSPKHQSPALQSTDDDEVIL 245
DB 733 SEFQVQAPARGTLGRVYPGSRSEKHSPPDSACSVDYSSCLSPSEH--PTDSESTEPLS 790
QY 246 VEGPVLQSSRLFTLKRCDRLVLRPMSEPLQNVVDHMANHLGVSPNRIILLGESE 305
DB 791 VDG-----TSSDLEEPAEGBEGGEGMGPP-----YGLQE 822
QY 306 LSPTATP-----STLKLGVADIIDCVWLASSEATETSQELRLRVQGEKHKQMLE 355
DB 823 GSP-QTPDQFQFLKQHPETLASGAPAP-VQVPERSESRISSRFLQLQVQTRPLREP-- 878
QY 356 ISLSPDSPLKVLMSHYBEAAGLSGHK 381
DB 879 ---SPSSSSSLALMSRPAQVPOASGEQ 901
RESULT 15
AAM79953

ID AAM79953 standard; Protein; 1301 AA.
XX
XX AAM79953;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 3599.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX WO200157190-A2.
PN
XX
XX 09-AUG-2001.
PD
XX 05-FEB-2001; 2001WO-US04098.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK53086.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 20; Page 393-394; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 1301 AA;
SQ
Query Match 7.4%; Score 154.5; DB 22; Length 1301;
Best Local Similarity 23.5%; Pred. No. 0.00043;
Matches 105; Conservative 48; Mismatches 164; Indels 129; Gaps 21;
QY 6 RRGGRSRGGRGARRGARGRCPRARQSPARLIPDTVLVDSDEEVLEADVPEVP 65
DB 537 RRGGRSQGFPSSQFQASG-----PNRHQAPSMLSFGPA---LSSDSKKEGEDEGTEELP 588
QY 66 -----VARLPAPAKPEQSDSDSEGAEE-----GPAGAPRTLVRRRRRLL 106
DB 589 ALPVLAKSTKKALASVSPALPRSLSHWMSRAQESVGFLLDPAPAAPNPGP---RRGRWV 645

```
Qy 107 DGEAFVVPVYSGKVOSLNLIPDNSLLKLCPS--EP--EDEADLTNGSGSPSEDDALPS 163
Db 646 QPG-----VELSVRSMDLRQLETLAPSLQDPQDSLAIIPSGPRKHGQEALET 694
Qy 164 G-----SPWRKKLRKKC-----EKEE---KKMEFPQDISPLPOPS---SRNK 201
Db 695 SLTSQNEKPPFPQASOPCSYPHIIRLLSOEBGVFAODLEPAPLEDGIVYEPBDNPTMDT 754
Qy 202 SRKHTALQK-----LREVNKRLQD-----LRSCLSPKCHOSPALQSTDEVVL 245
Db 755 SEFOQAAPARGLGRVYPSGRSEKHSPPSACSVDYSSCLSSPEH--PTEDSESTEPLS 812
Qy 246 VEGPVLPOSSRLFTLKIRCRADLVRLPVRMSEPLQNVVDHMANHGVSPNRILLFGESE 305
Db 813 VDG-----ISSDLBEPAEGDEEEEGGMP-----YGLQE 844
Qy 306 LSPTATP-----STLKIQVADIIDCVVLASSSEATETSOELRLRVQGEKHKQMLE 355
Db 845 GSP-OTPDOEQFLKOHFTETLAGAAPGAP-VQVPERSESRSSISRPFLQVQTRPLREP-- 900
Qy 356 ISLSPDSPLKVLMSHYEAMGLSGHK 381
Db 901 ---SPSSSSLALMSRPAQVPQASGEQ 923
```

Search completed: January 12, 2003, 10:26:29
Job time : 42 secs

This Page Blank (copy)